将棋名人のレーティングと棋譜分析

山下 宏

2014年11月7日

GPW 箱根

大山15世名人と羽生名人、全盛期に戦えばどちらが強い？

- 大山康晴15世名人
  - タイトル獲得80期
  - 昭和の覇者
- 羽生善治名人
  - 1996年に7冠達成
  - 平成の覇者

歴代名人の強さを調べる

- 対局の結果から
  - 対局者の棋力を点数で表す
  - 勝てば点数プラス、負ければマイナス
  - いわゆるEloレーティング
- 棋譜の内容から
  - 棋譜を将棋プログラムで解析して強さを推定
  - 勝敗の情報は使わない
  - こちらがメインです

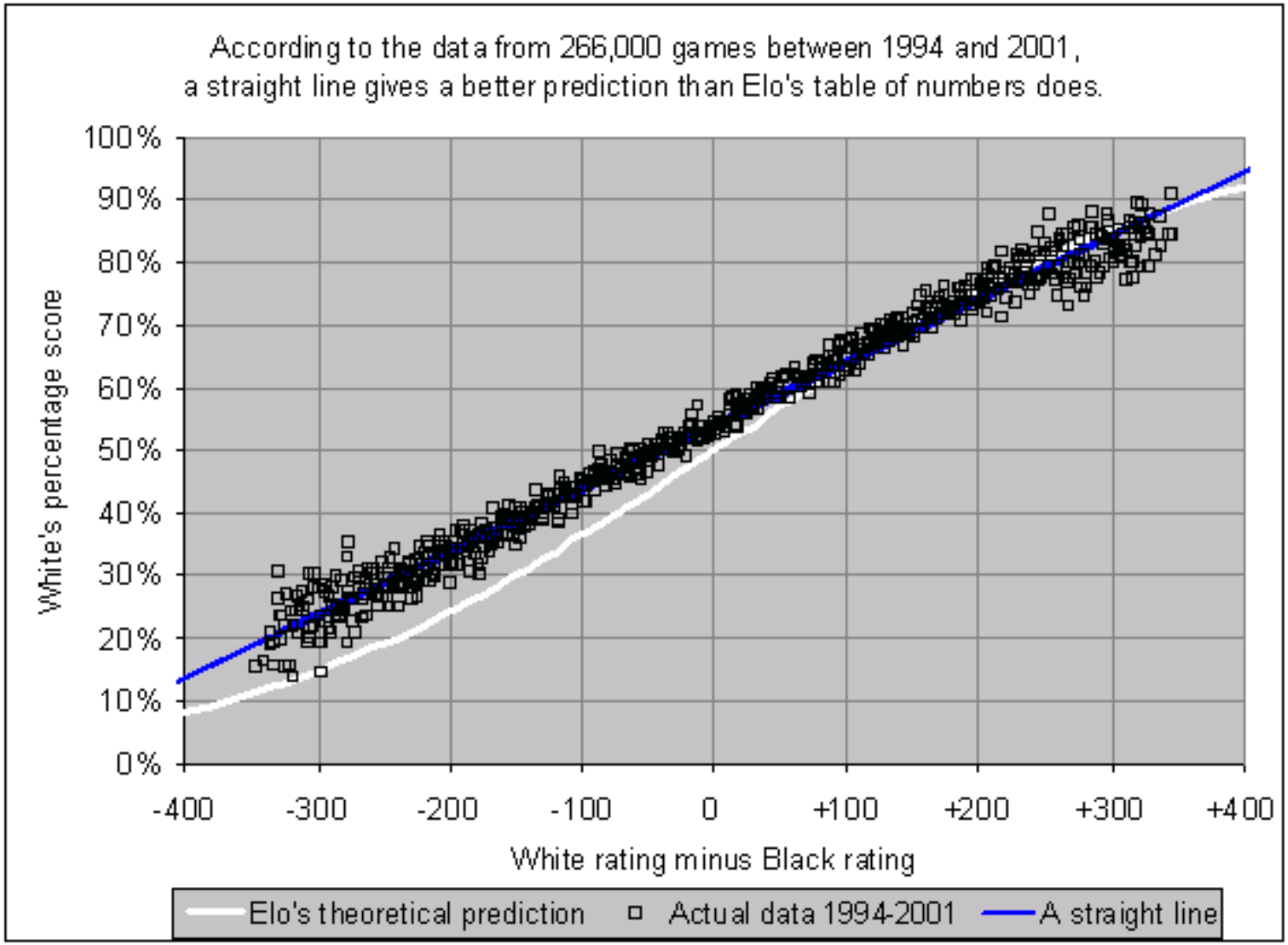
棋譜

- 棋譜でーたーべーす
  - ネットでの匿名による棋譜の登録
  - 56，000局
  - プロがメインだがアマや女流も含む
  - 重複や入力ミスがある
    - 日付が同じ棋譜だけは修正
- 将棋倶楽部24の24万局集
  - 対局サイトの棋譜からアマの棋譜を24万局
  - 2004年出版。今は絶版。

対局の結果からのレーティング

- Eloレーティング
  - 対局者の強さを点数で表す
  - 100点差で勝率64%、200点差で75%になる
    - 同じ点数同士。勝者+16点、敗者−16点
  - 200点差。強い方が勝ち
    - 勝者+8点、敗者−8点
  - 200点差。弱い方が勝ち
    - 勝者+24点、敗者−24点
  - 3勝1敗ペース（75%）で点数の変動なし

実際は理論勝率からずれるらしい

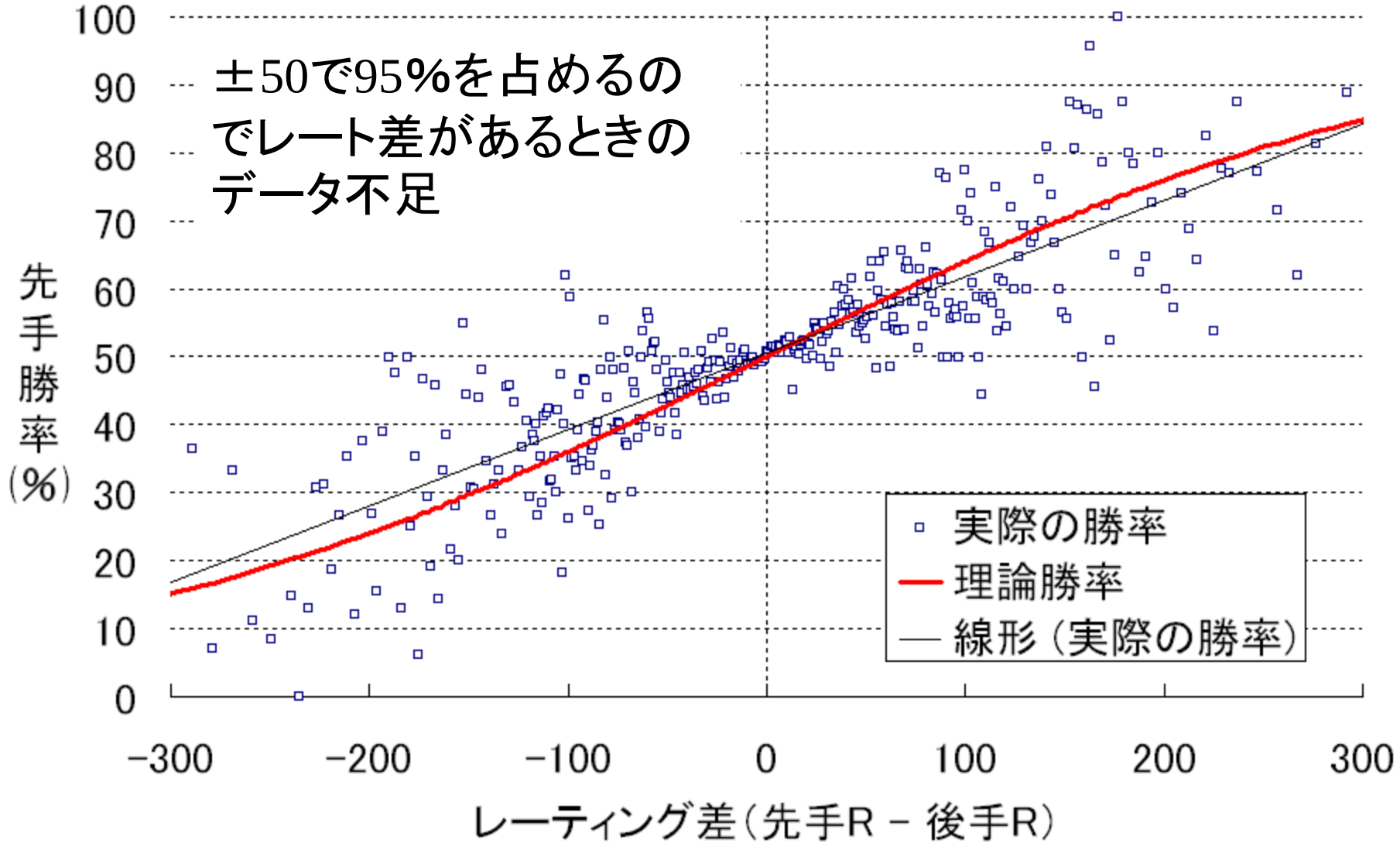


チェスの場合

白番の勝率が互角で54%近い

The Sonas Rating Formulaから引用

将棋倶楽部24のレーティング差と勝率



Eloの計算式

$$E_A = \frac{1}{1 + 10^{(R_B - R_A)/400}}$$

$$R'_A = R_A + K(S_A - E_A)$$

R_A : 自分のレーティング
R_B : 相手のレーティング
R'_A : 自分の新しいレーティング
S_A : 勝ちで 1、負けで 0
K : 変動率, $K = 16$ を用いた

Eloレーティングの特徴

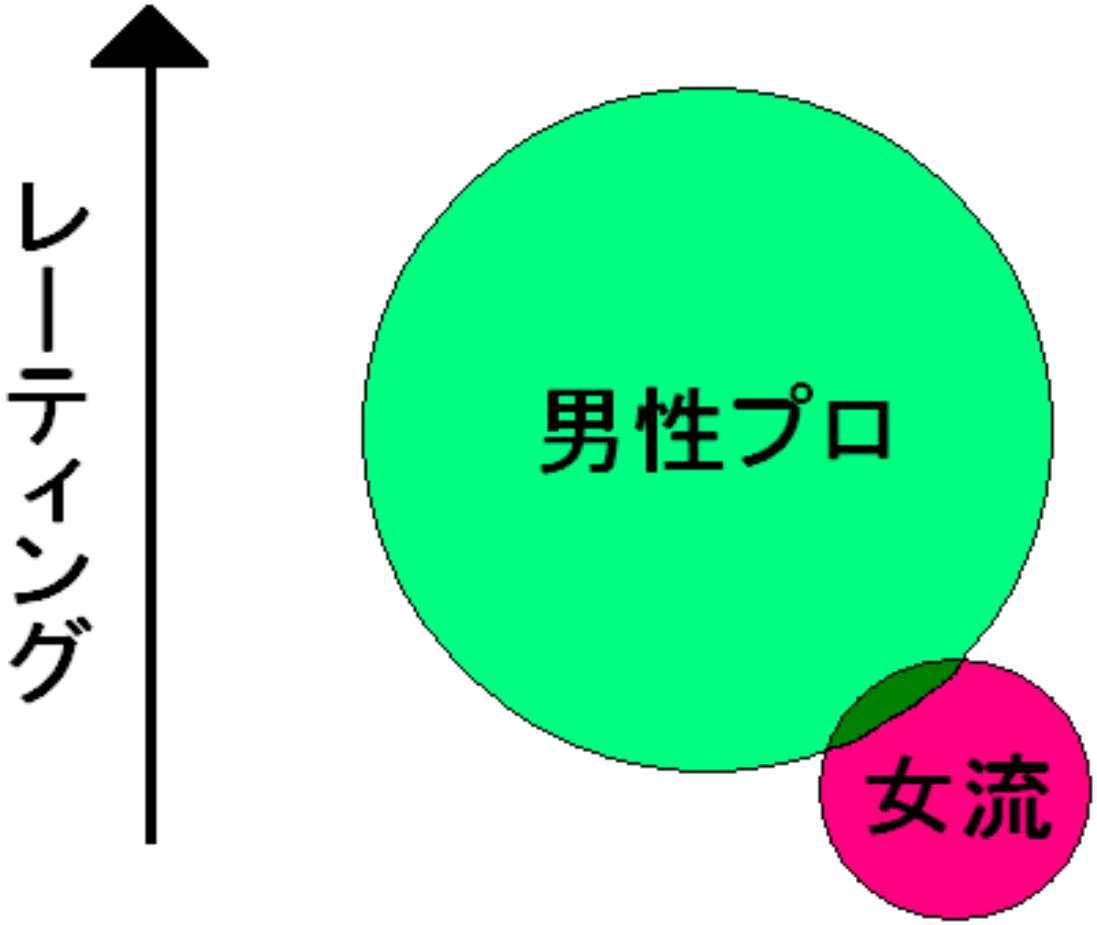
- 点数の差のみに意味がある
- 点数の大きさに意味はない
  - 全員に＋500点、など任意の数字を足せる
- 計算が簡単
  - 将棋倶楽部24はEloの簡易版を利用

プロ棋士のEloレーティングを計算

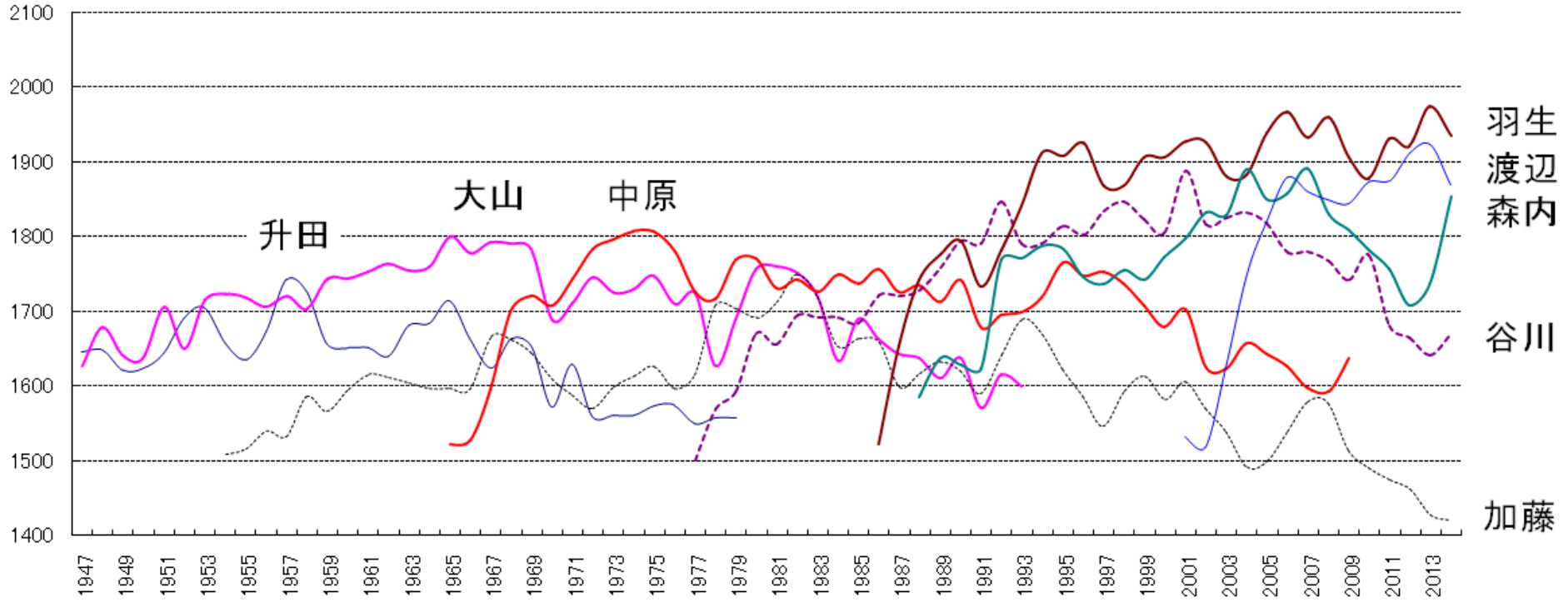
- 全員の初期値を1500点で計算
  - 清水市代さんが1578点に
    - 高すぎる！
  - 男性に163局で勝率0.18
  - 女流に523局で勝率0.65
  - 女流を除くと1286点まで下がる
- 女流を含めた計算で清水さんが1286点になるように女流の初期値を1074点にした

Eloレーティングの弱点

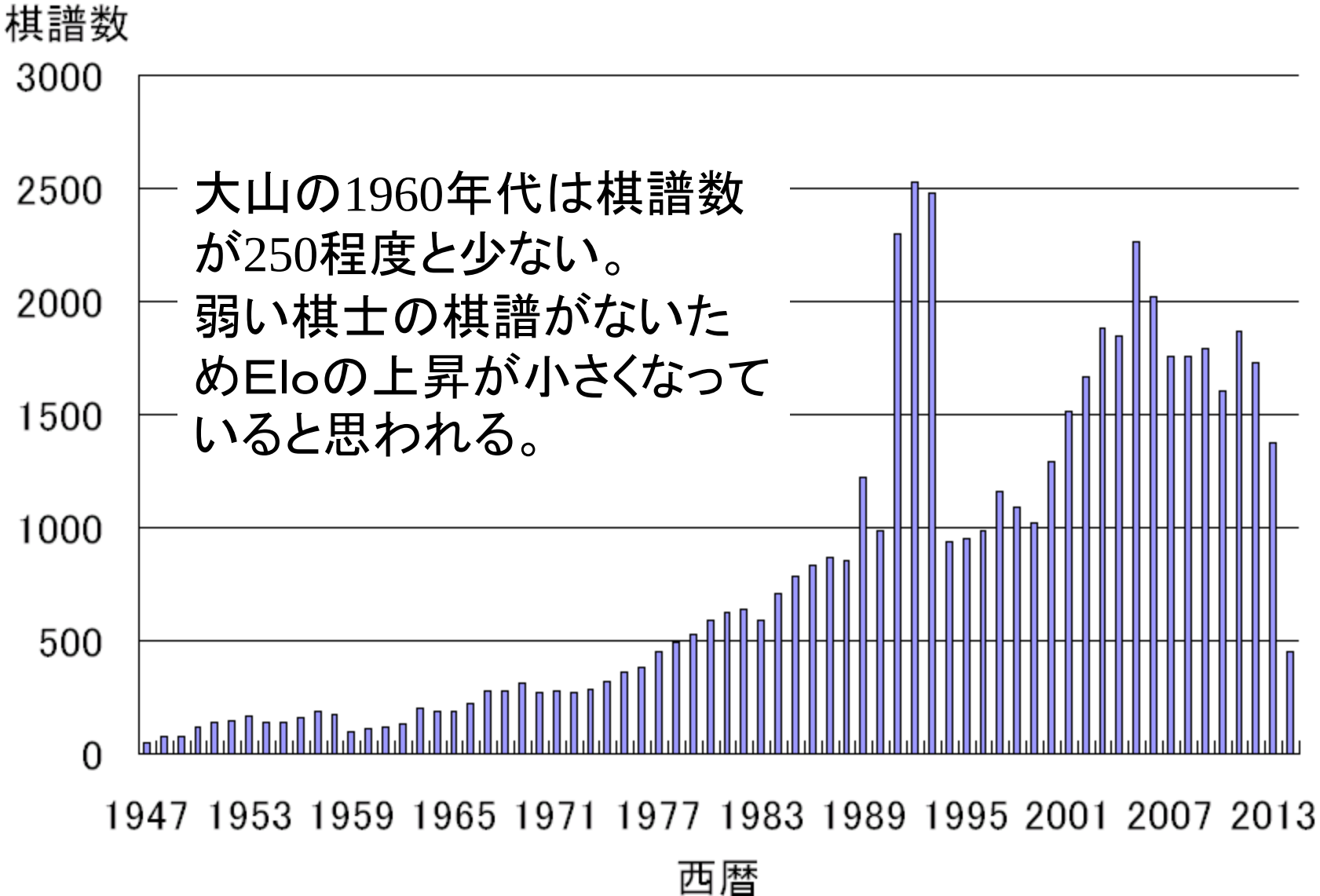
母集団が異なり、相互の対局数が少ないときは苦手



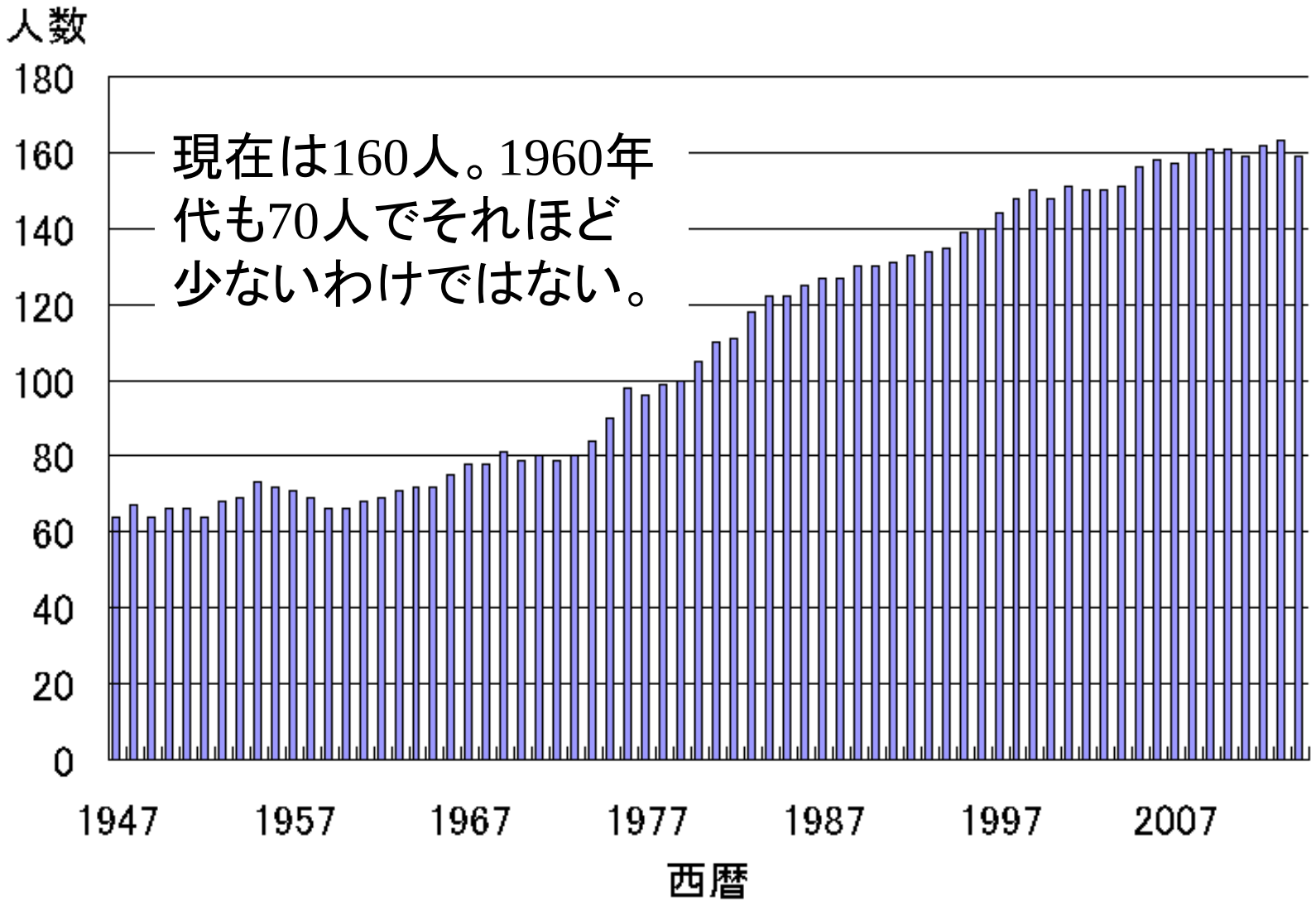
歴代名人のEloレーティングの推移



「棋譜でーたべーす」の年毎の棋譜数



現役プロ棋士の人数の推移



Whole－History Rating（WHR）

- Eloが1局のみの結果で更新するのに対し、WHRはすべての対局結果から最大事後確率を計算する
- レーティングの高い方が勝つ、という予想が
  - Elo　　55.1% 当たる
  - WHR　55.8% 当たる
- わずかだがWHRの方が予測性能がいい

WHRの計算例 1

- 例えば下記の結果があった場合

3人のプレイヤ（A、B、C）がいて

1日目　A が B に勝つ
2日目　A が B に勝つ
3日目　C が A に勝つ

WHRの計算例 2

- 1日目（AがBに勝つ）で計算

	A	B	C
1日目	+92	-92	---

WHRの計算例 3

- 2日目（AがBにまた勝つ）までで計算

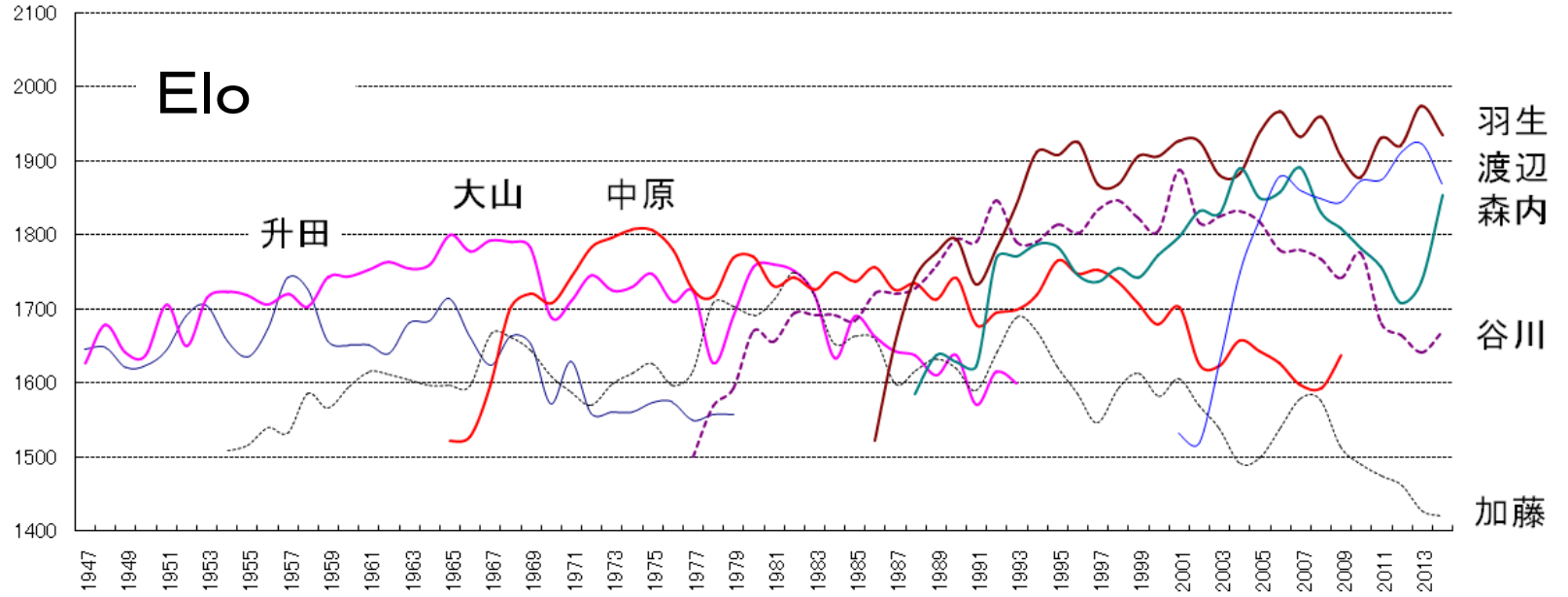
	A	B	C
1日目	+131	-131	---
2日目	+132	-132	---

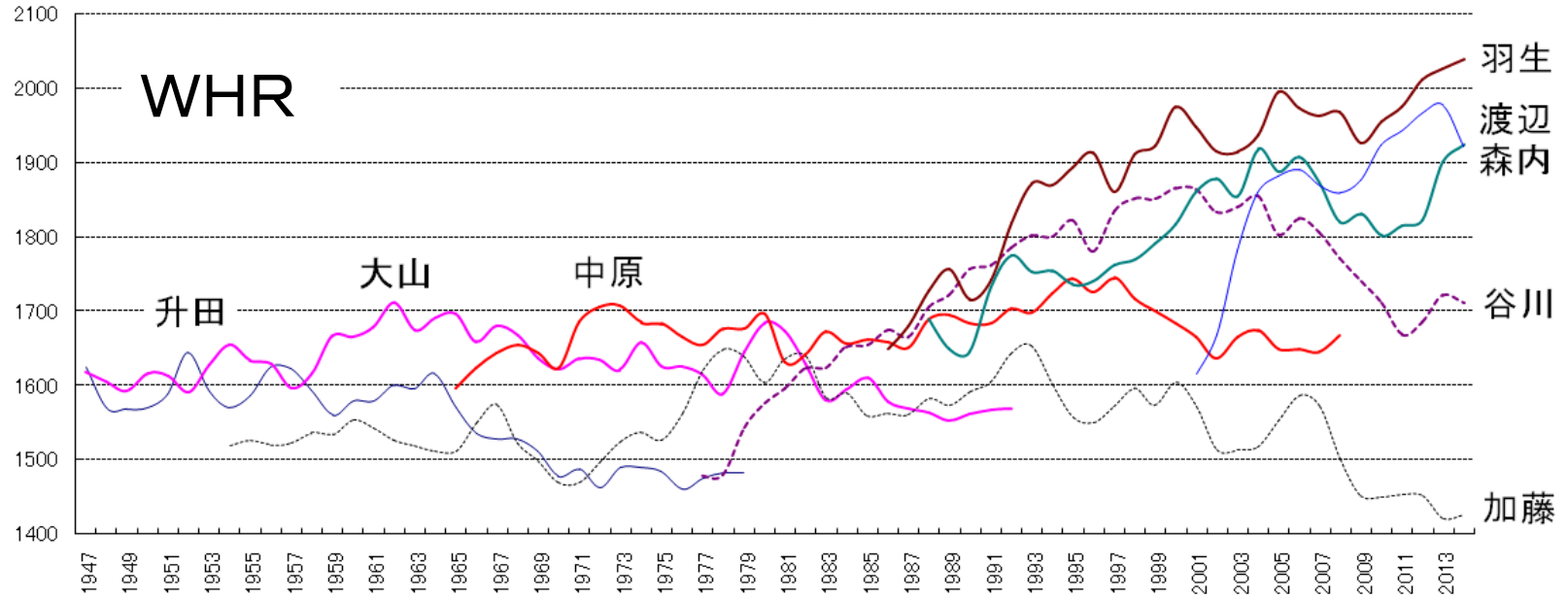
WHRの計算例 4

- 3日目（CがAに勝つ）までで計算

	A	B	C
1日目	+35	-174	---
2日目	+35	-174	---
3日目	+34	-174	+132

3日目は対局していない B の点も下がり、同時に A, B の過去の点も変化する


Elo
2100
2000
1900
1800
1700
1600
1500
1400
1947
1949
1951
1953
1955
1957
1959
1961
1963
1965
1967
1969
1971
1973
1975
1977
1979
1981
1983
1985
1987
1989
1991
1993
1995
1997
1999
2001
2003
2005
2007
2009
2011
2013
升田
大山
中原
羽生
渡辺
森内
谷川
加藤


WHR
2100
2000
1900
1800
1700
1600
1500
1400
1947
1949
1951
1953
1955
1957
1959
1961
1963
1965
1967
1969
1971
1973
1975
1977
1979
1981
1983
1985
1987
1989
1991
1993
1995
1997
1999
2001
2003
2005
2007
2009
2011
2013
升田
大山
中原
羽生
渡辺
森内
谷川
加藤

棋譜の内容からレーティングを計算

- Bonanza6.0を使って棋譜の1手1手を探索
  - 最善手と評価値を記録
- 探索深さは11に固定
- 詰をBonanzaのdfpn、1000万ノードで調べる

棋力との関連性が高い指標を見つけたい

- 「平均悪手」という指標を導入する
- Bonanzaと違う手を指して、かつ評価が下がったときを悪手とする

平均悪手の計算例

- ▲7六歩をBonanzaは選び評価は +0.3
- 人間は▲2六歩を指した
  - 次の▲2六歩の局面の評価は +0.2
- 評価が下がっているので悪手と認定
  - 「悪手合計」+= 0.1
  - 「平均悪手」＝「悪手合計」÷「手数」

平均悪手で無視する手

- 40手目以降のみを対象
  - 定跡の排除
- +10以上の評価は無視
  - 「形作り」の手を排除
  - 詰みが絡む局面が必要以上に影響するのを避ける

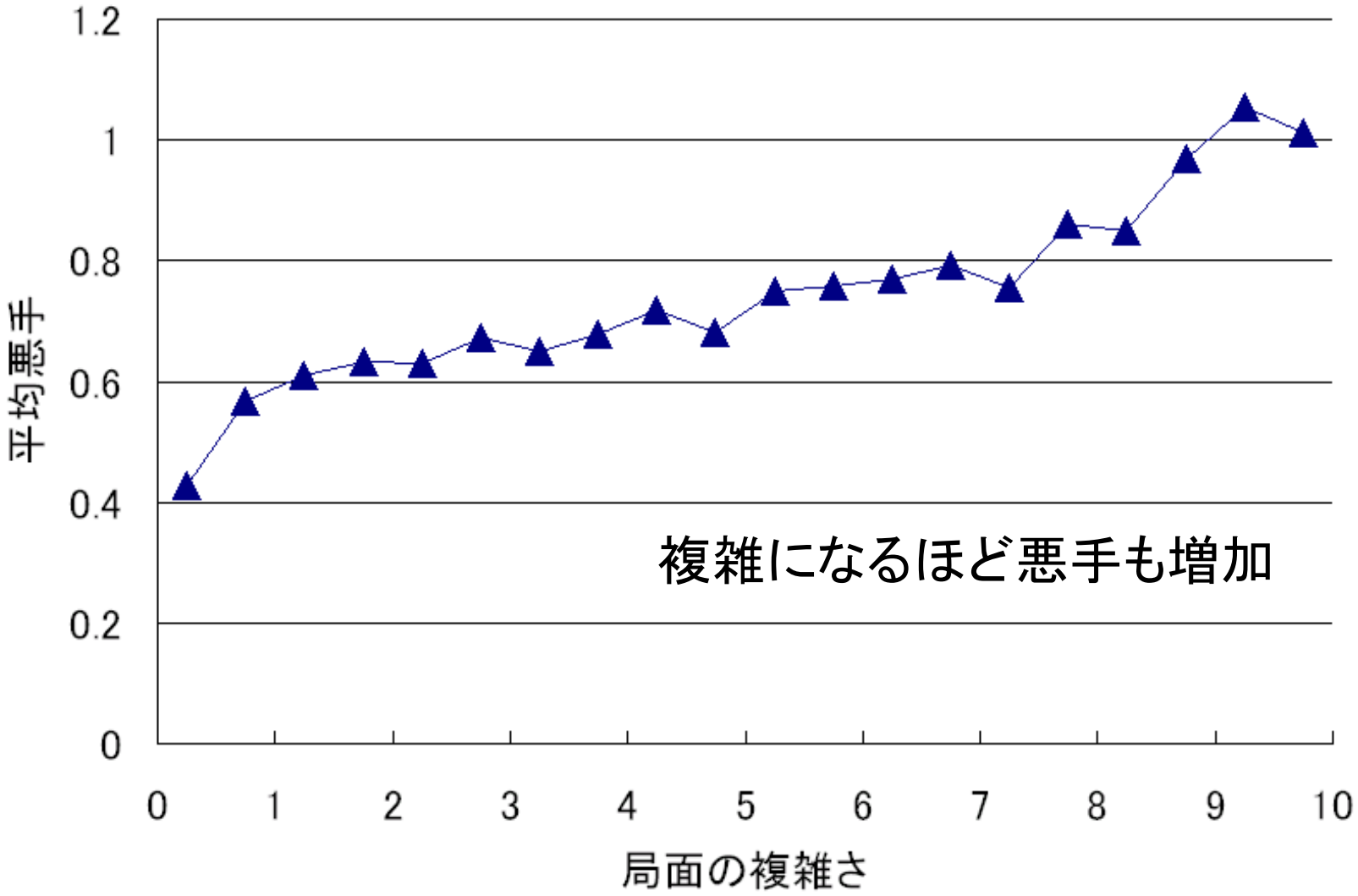
同様に

- 「平均好手」
  - Bonanzaと違う手を指して評価が上がった場合
- 「一致率」
  - Bonanzaと同じ手を指した割合
- 「好手率」
  - 好手を指した割合

「複雑さ」

- 局面の複雑さを表す指標
- 反復深化の途中で最善手が変わった場合の評価値の差の合計
  - 深さ1　▲76歩 +0.3
  - 深さ2　▲26歩 +0.2
  - 深さ3　▲26歩 +0.0
  - 深さ 1 → 2 のみが対象で、複雑さは 0.1
- 「平均変動」という名称が正しいかも

「複雑さ」と「平均悪手」の関係

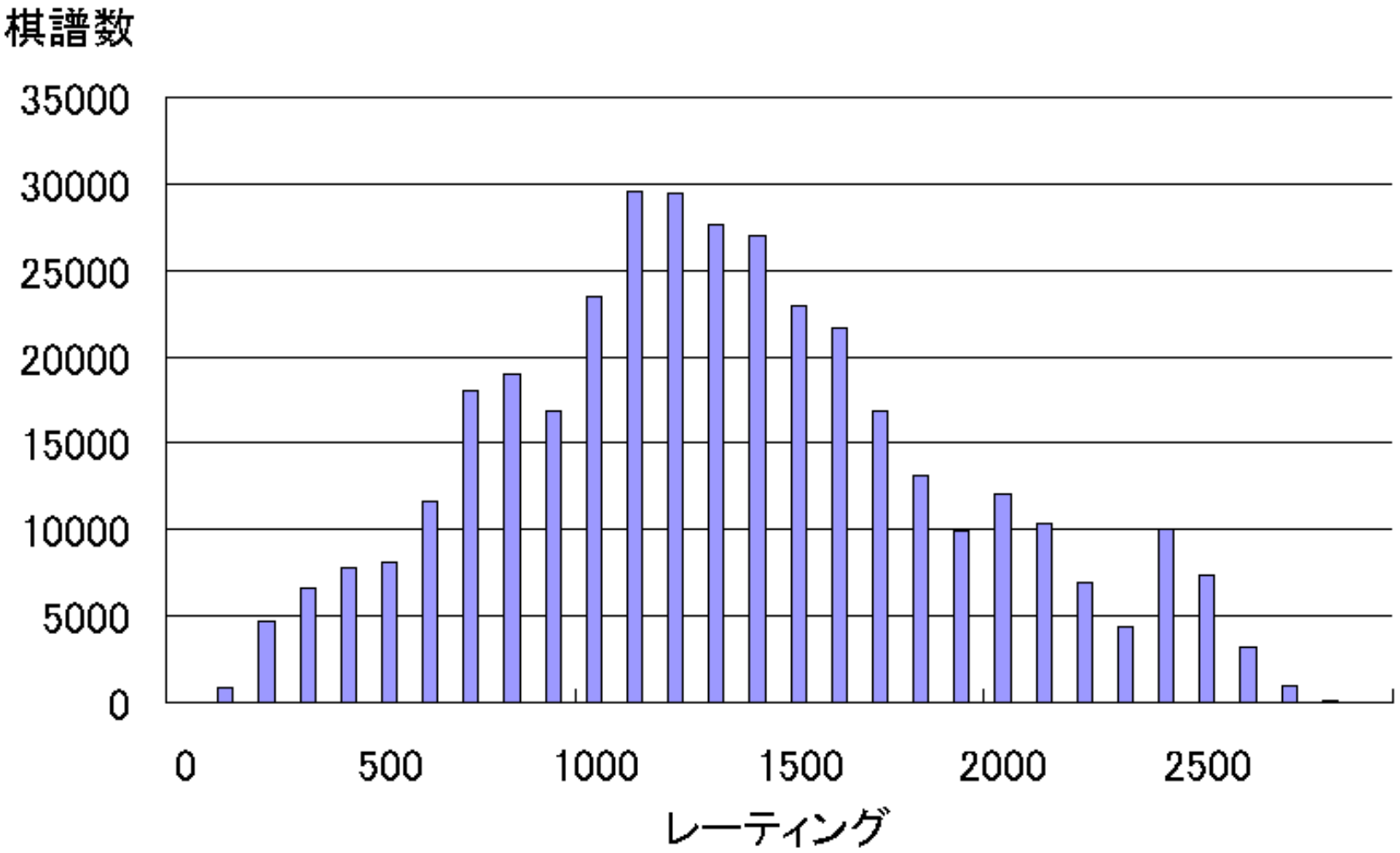


将棋倶楽部２４の棋譜を解析

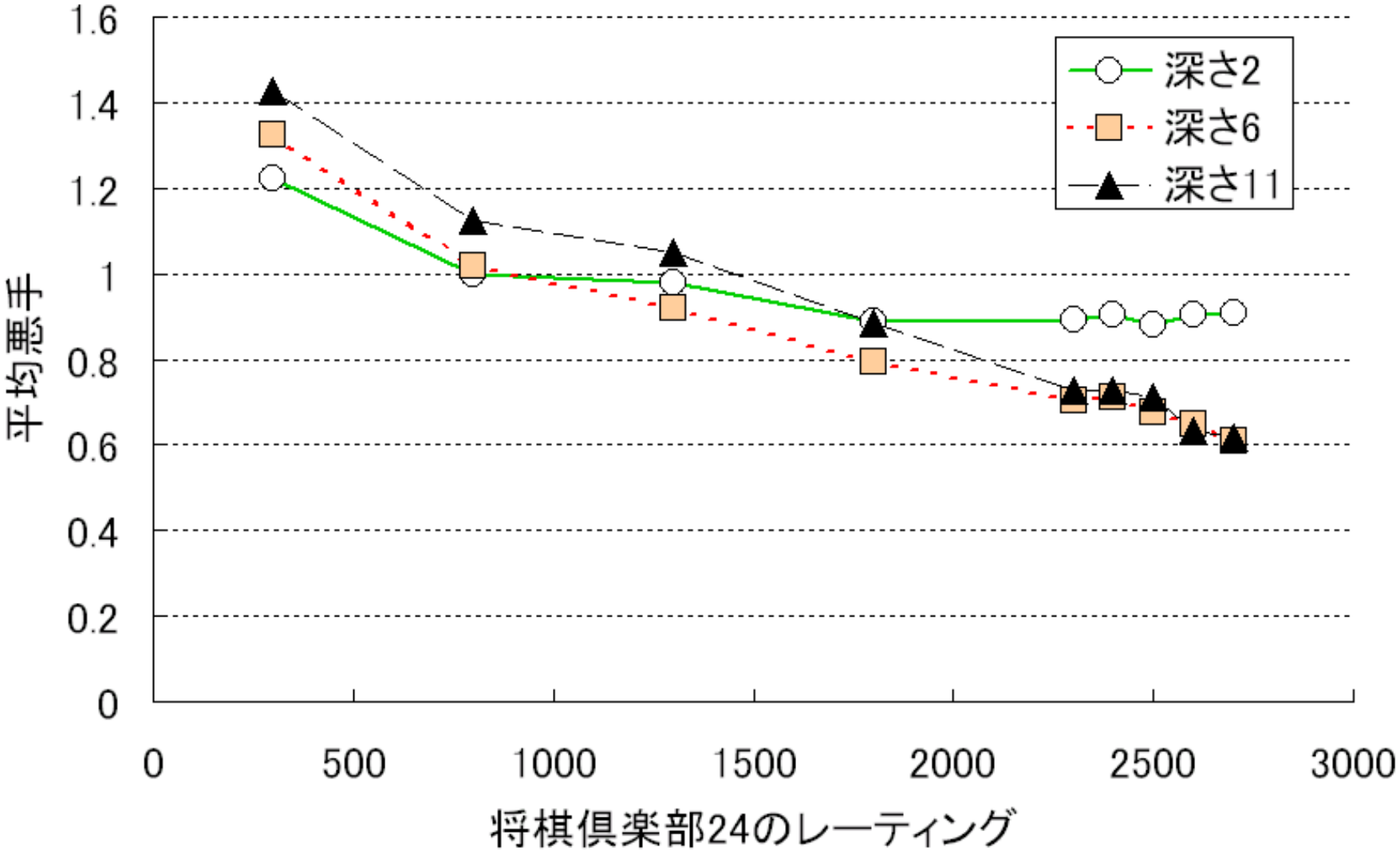
- R300、R800、R1300、R1800、R2300、R2400、R2500、R2600、R2700
  の棋譜を200局ずつ調べた
- R300の棋譜は R350 ±30 同士の対戦

- Bonanzaの深さは１１だけでなく、
  深さ2、深さ6も調べた

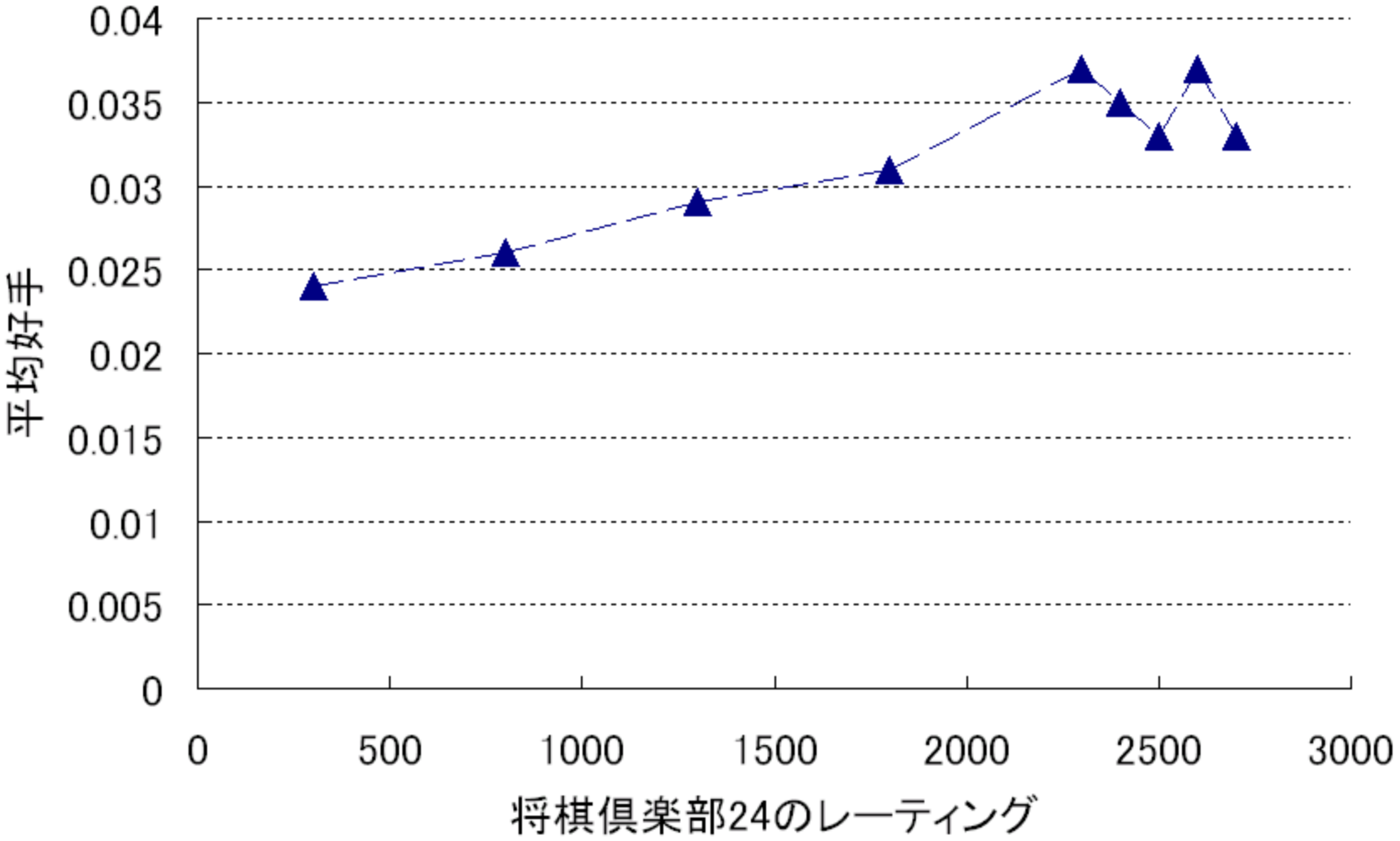
将棋倶楽部24のレーティング分布



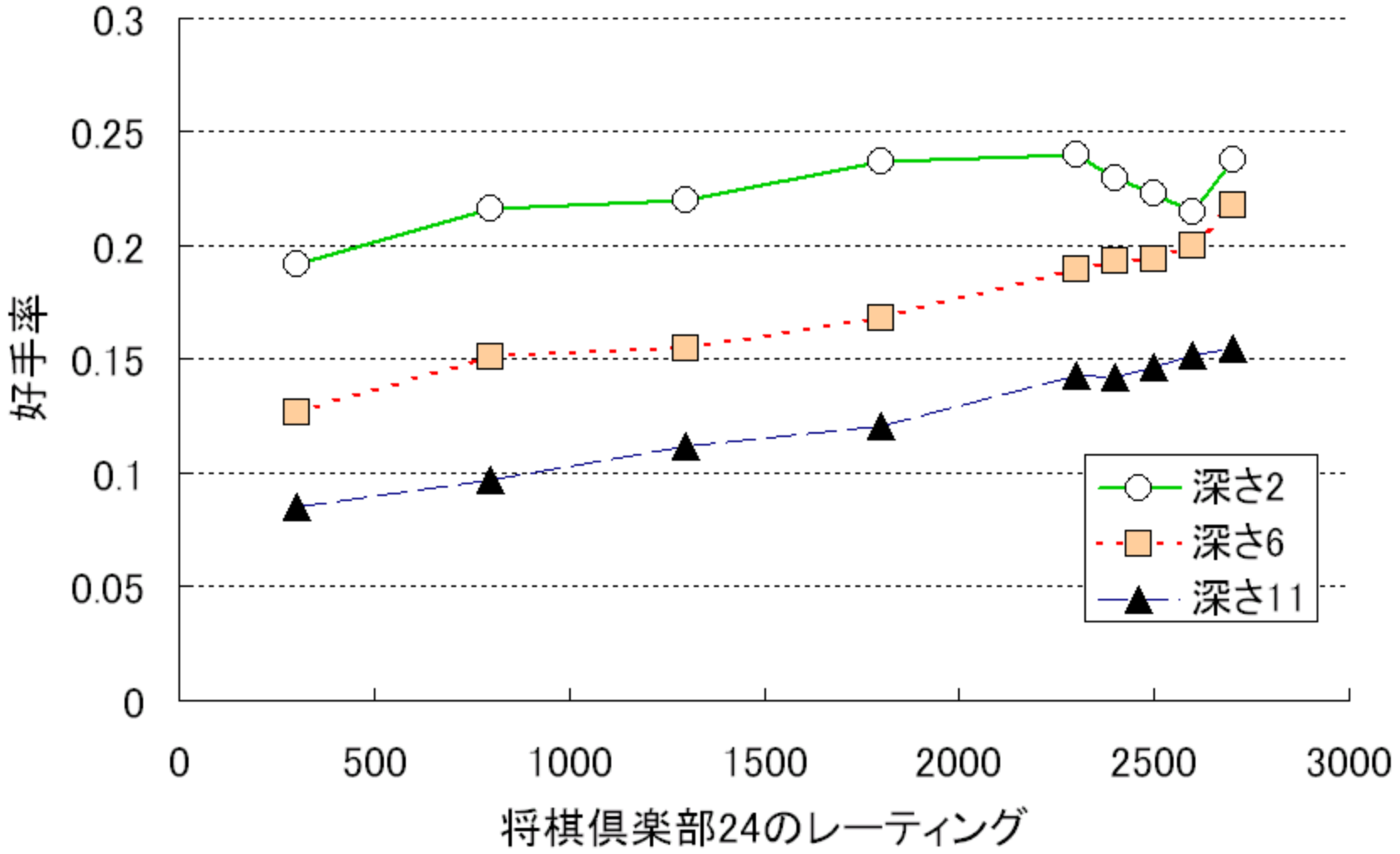
深さ2、6、11での平均悪手



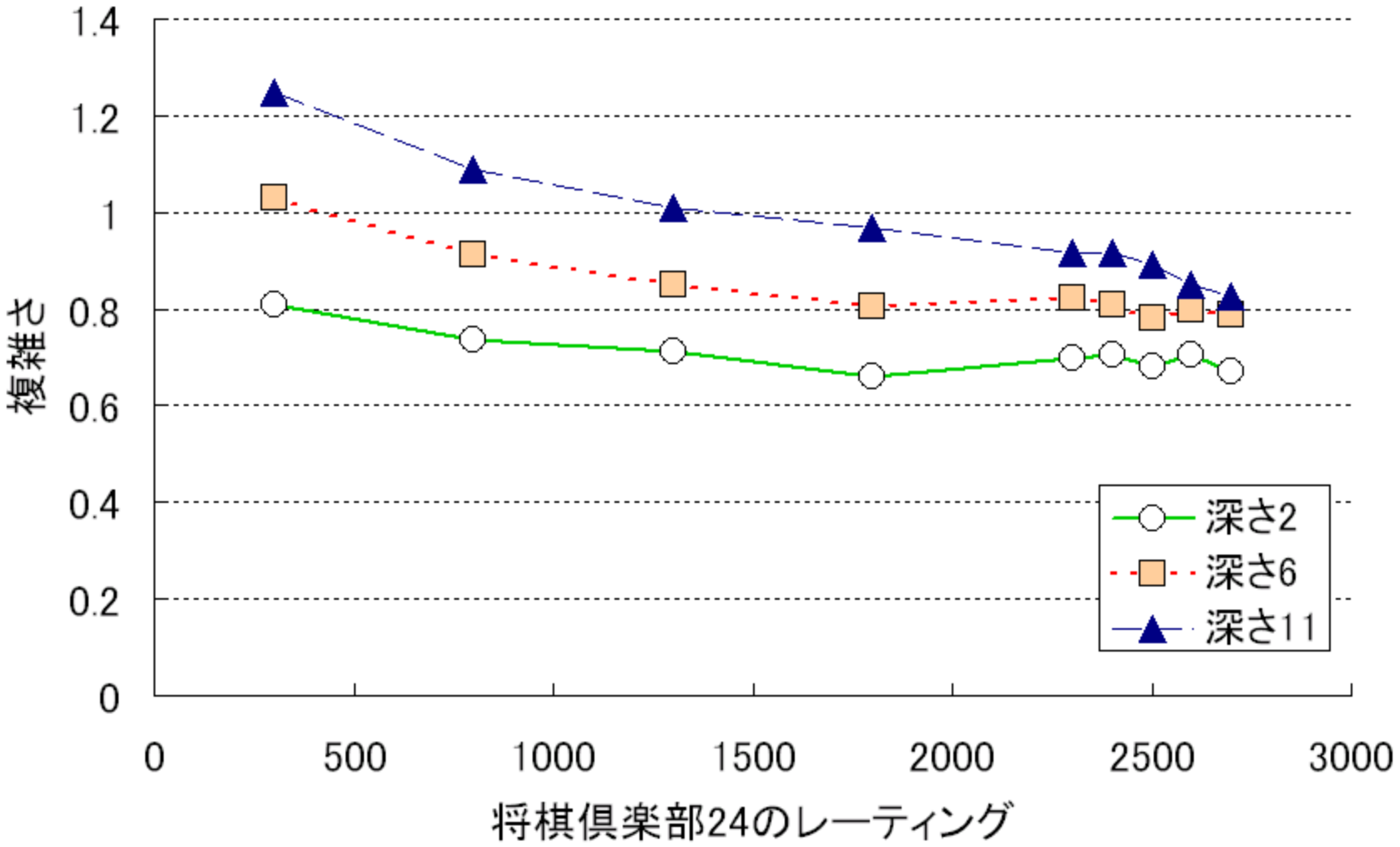
深さ11での平均好手


0.04
0.035
0.03
0.025
0.02
0.015
0.01
0.005
0
平均好手
0
500
1000
1500
2000
2500
3000
将棋倶楽部24のレーティング

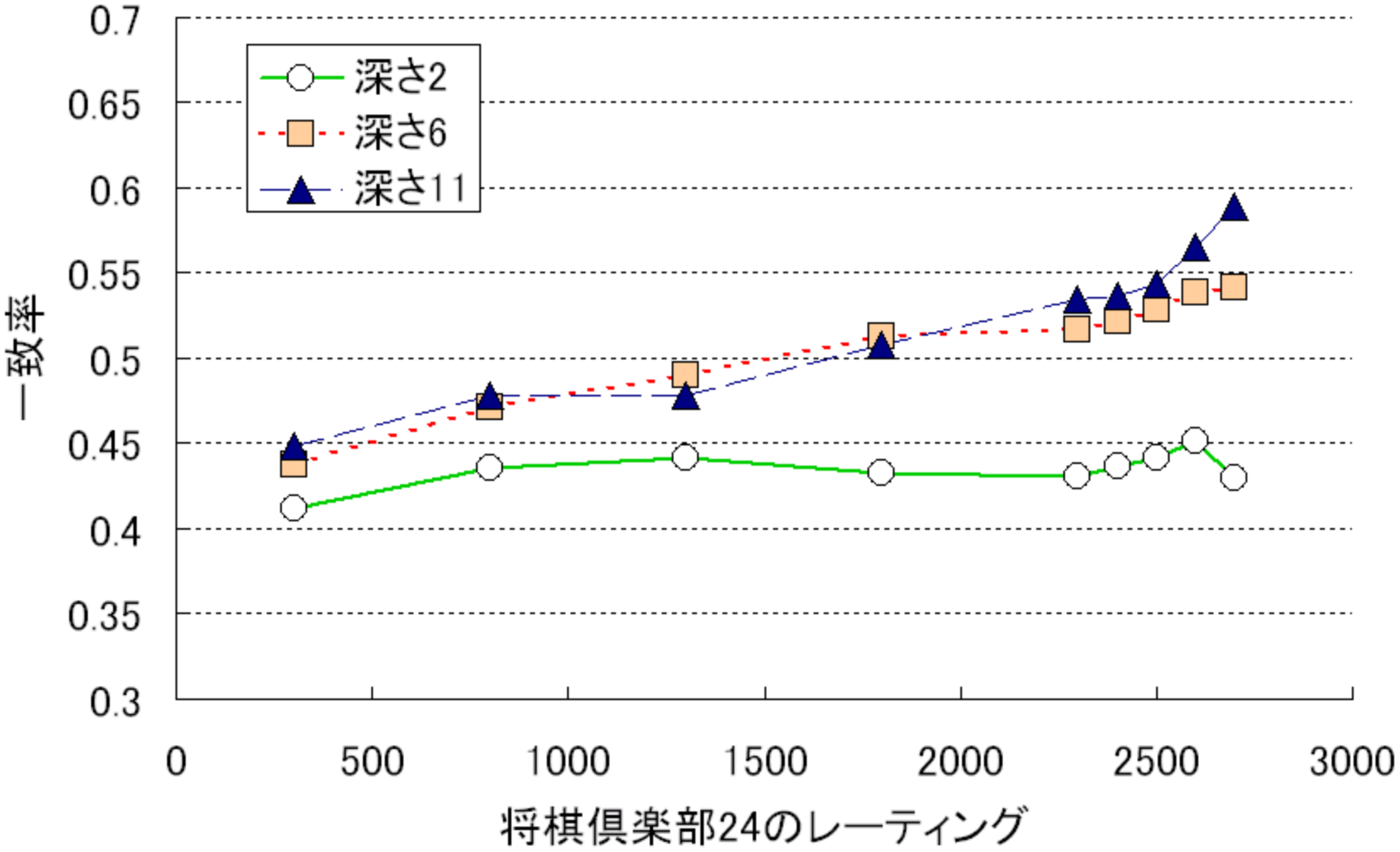
深さ2、6、11での好手率



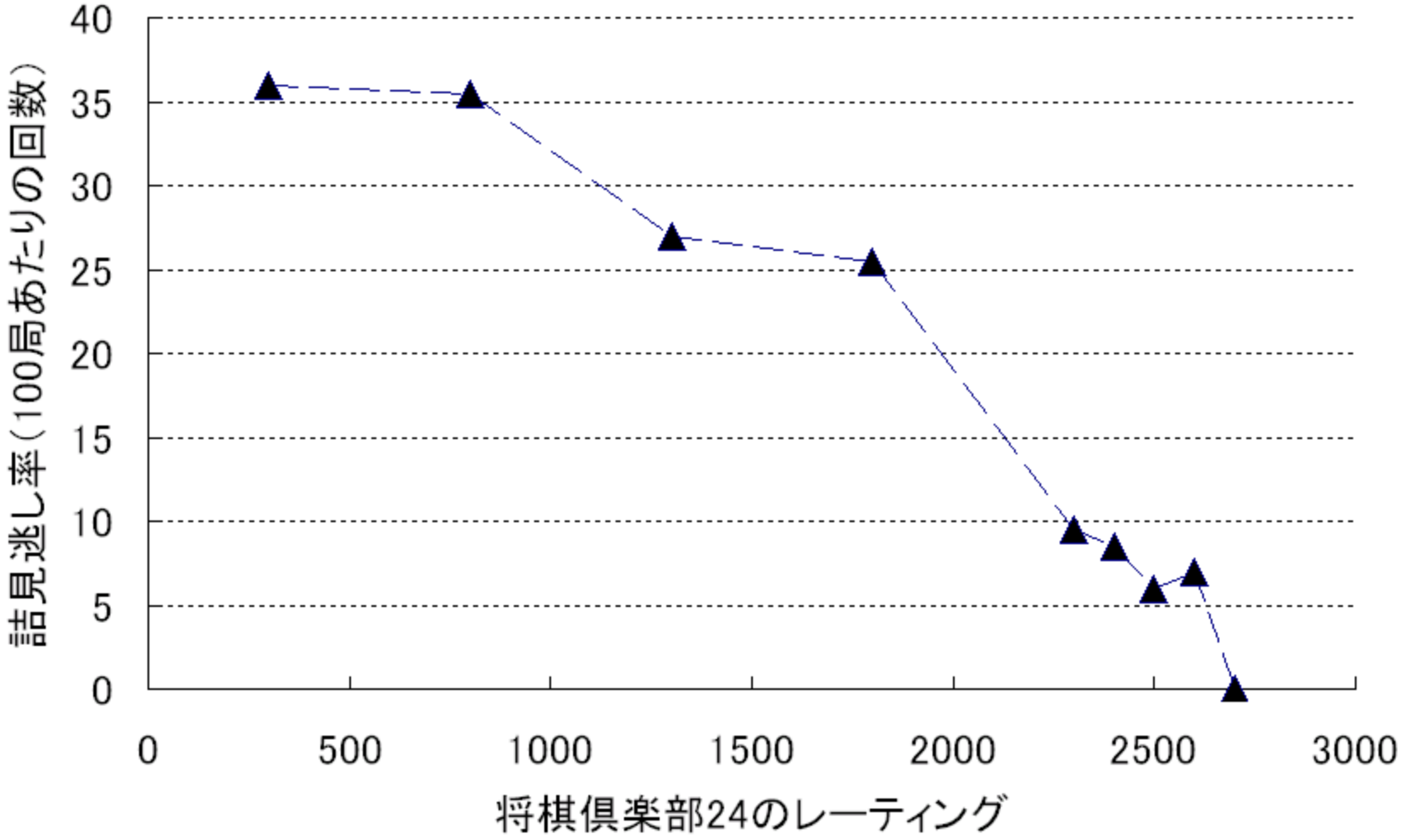
深さ2、6、11での複雑さ



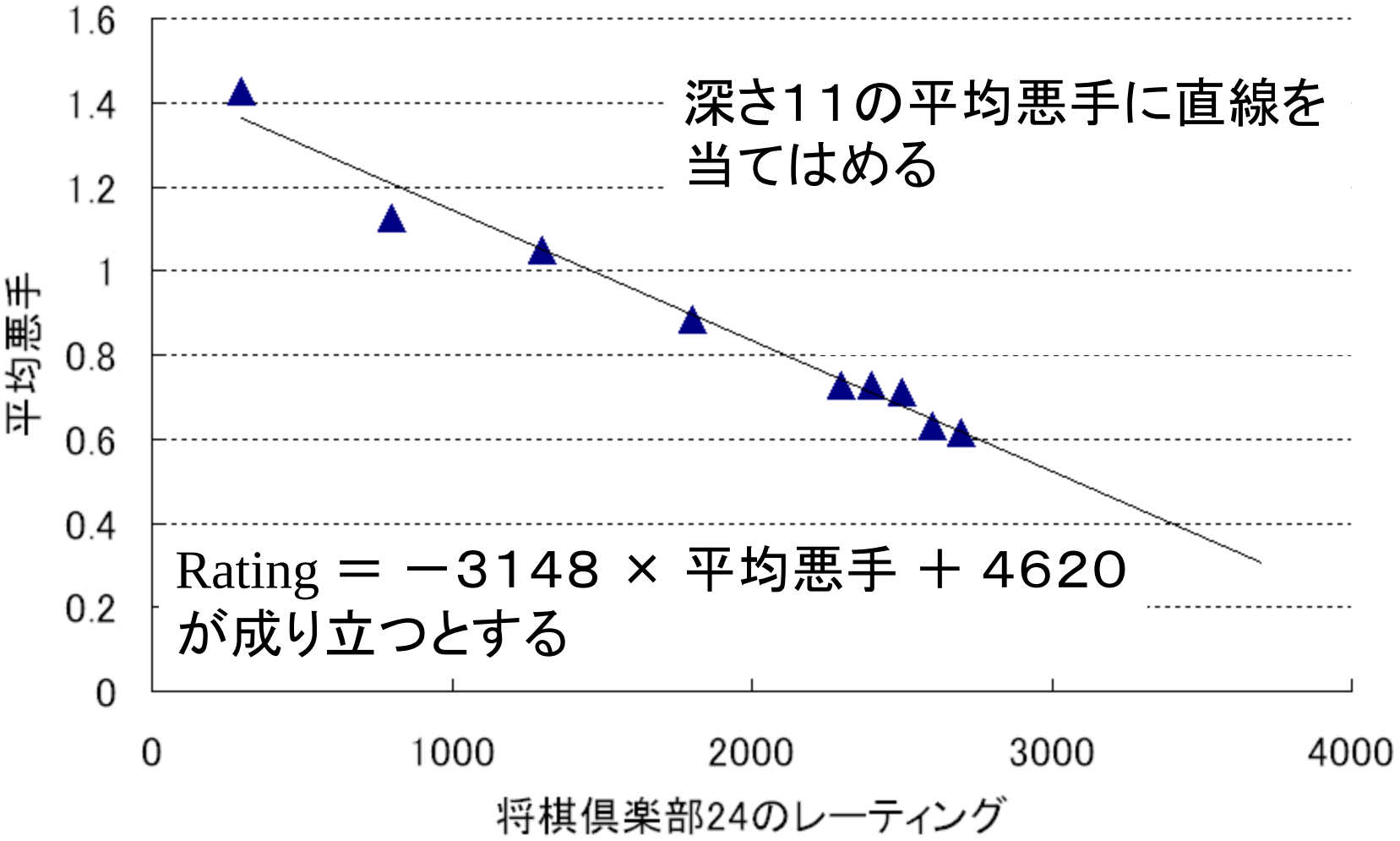
深さ2、6、11での一致率



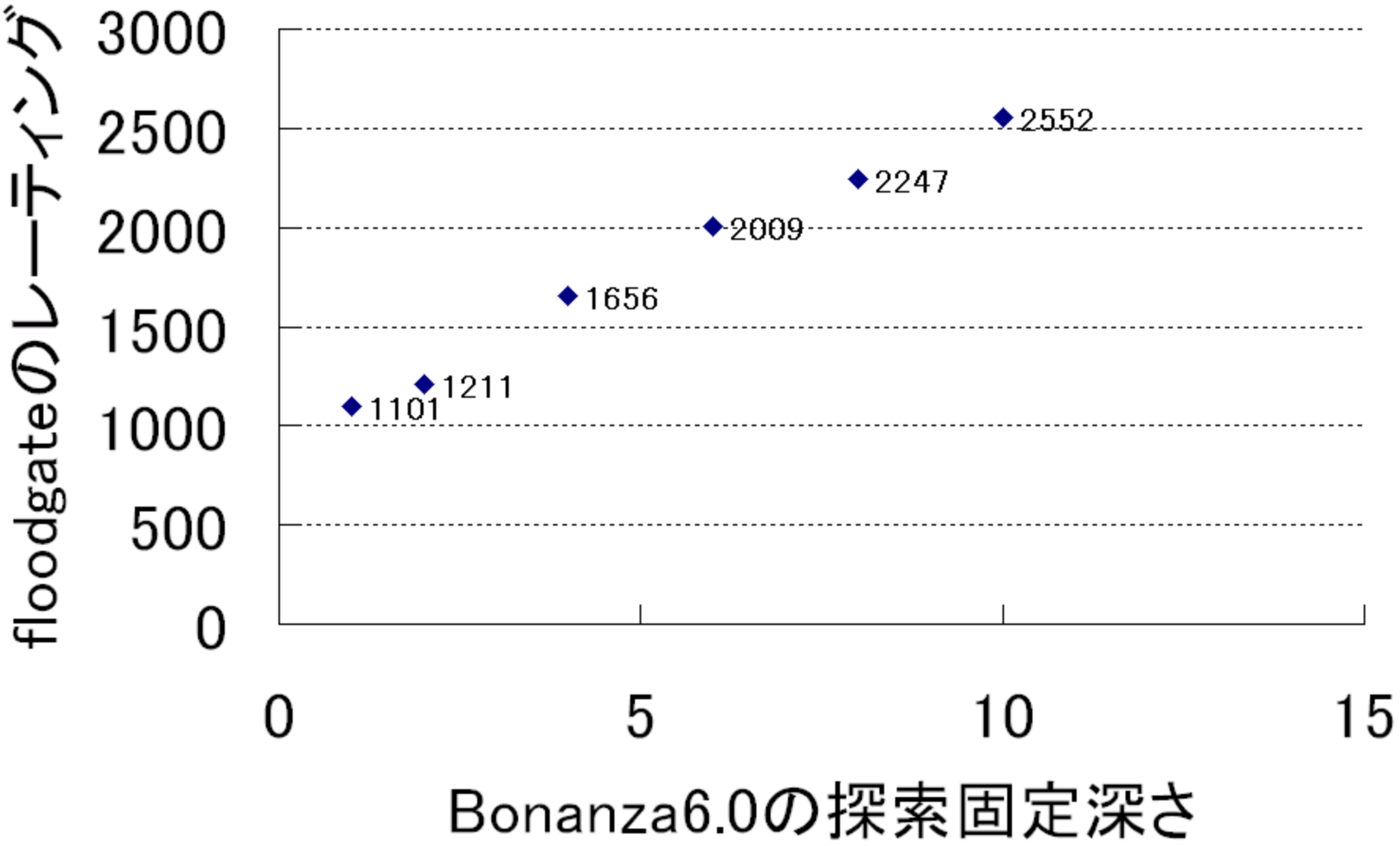
詰見逃し率（100局あたりの回数）



平均悪手がもっとも予測できそう



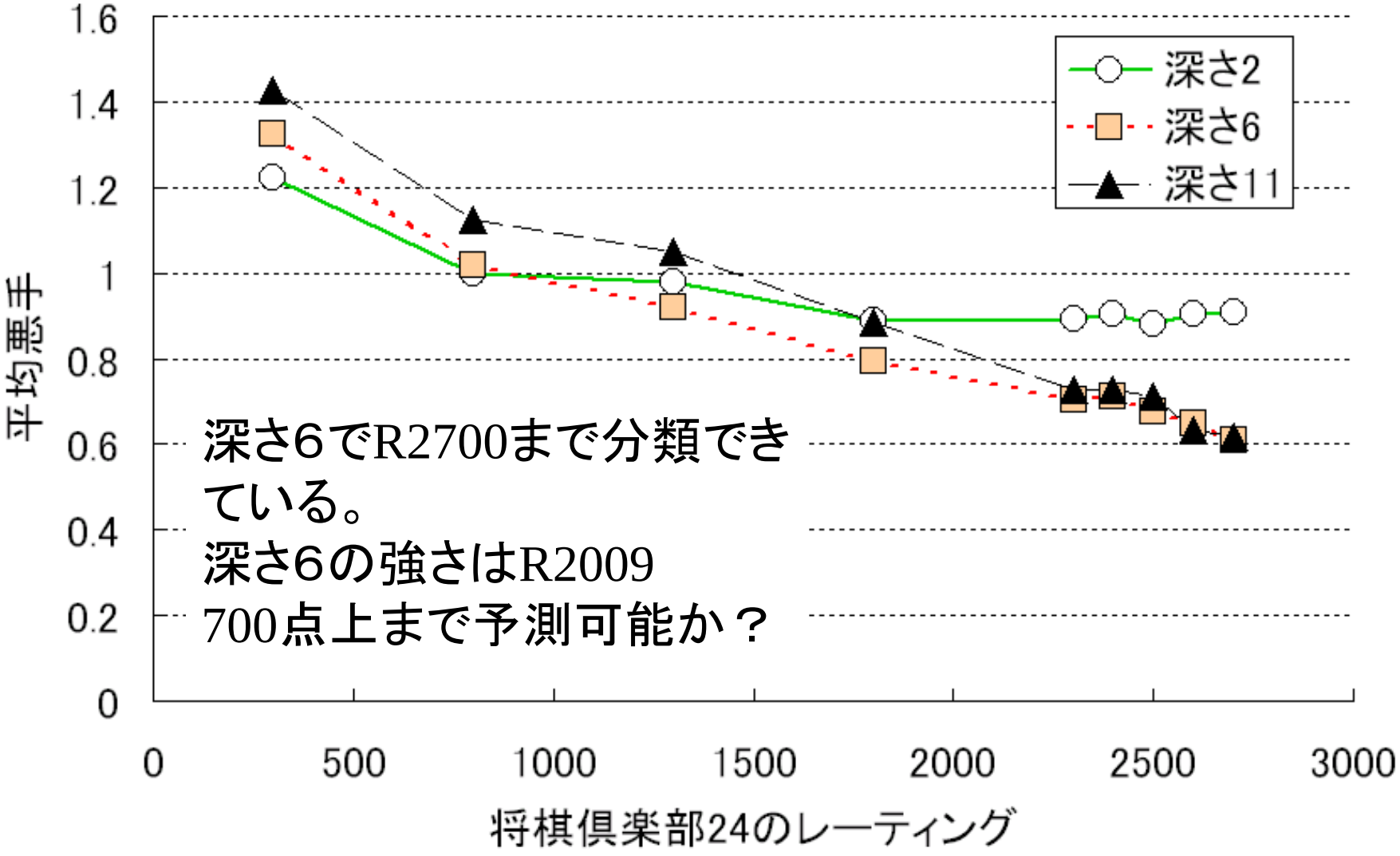
解析に使ったBonanzaの強さ



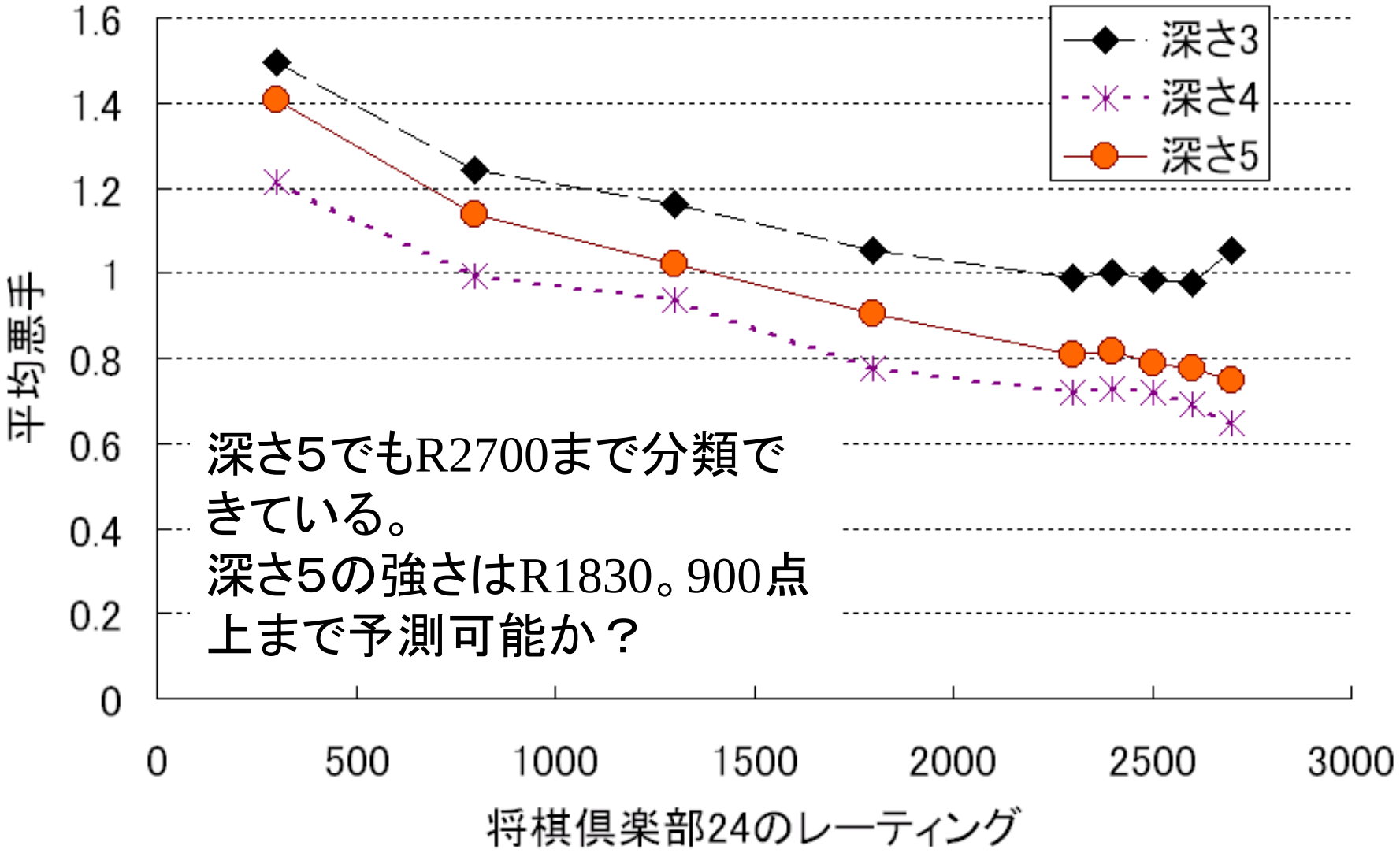
floodgateと将棋倶楽部24のレーティングを直接比較できるとする

- 2004年の将棋倶楽部24との比較
  - 将棋倶楽部24は年々インフレしている？
- floodgateは2007年の将棋倶楽部24のYSSの点数と一致するように調整された
- ちょっと乱暴だがそれほど差はないはず

深さ2、6、11での平均悪手（再掲）



深さ3、4、5での平均悪手



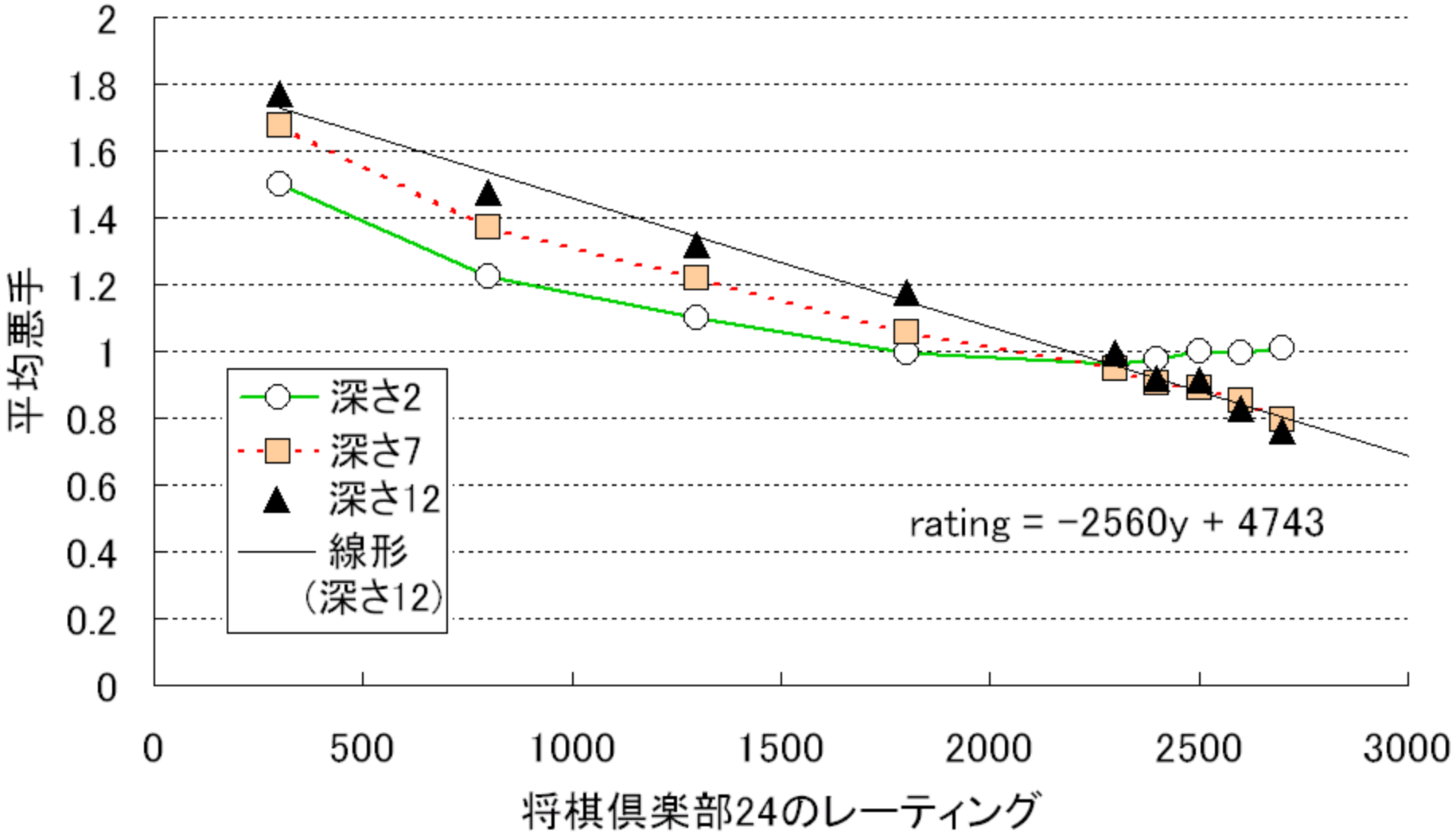
この手法の欠陥

- プログラムより強い棋譜は分類できないはず
  - 深さ2（R1200）でもR1800まで分類できている？
  - 自分より強い棋譜も分類可能か
- プログラム自身の棋譜を解析すれば平均悪手は0になってしまう
  - 明らかに正しくない！
- しかし人間の棋譜に関しては強い関連性があるように思われる

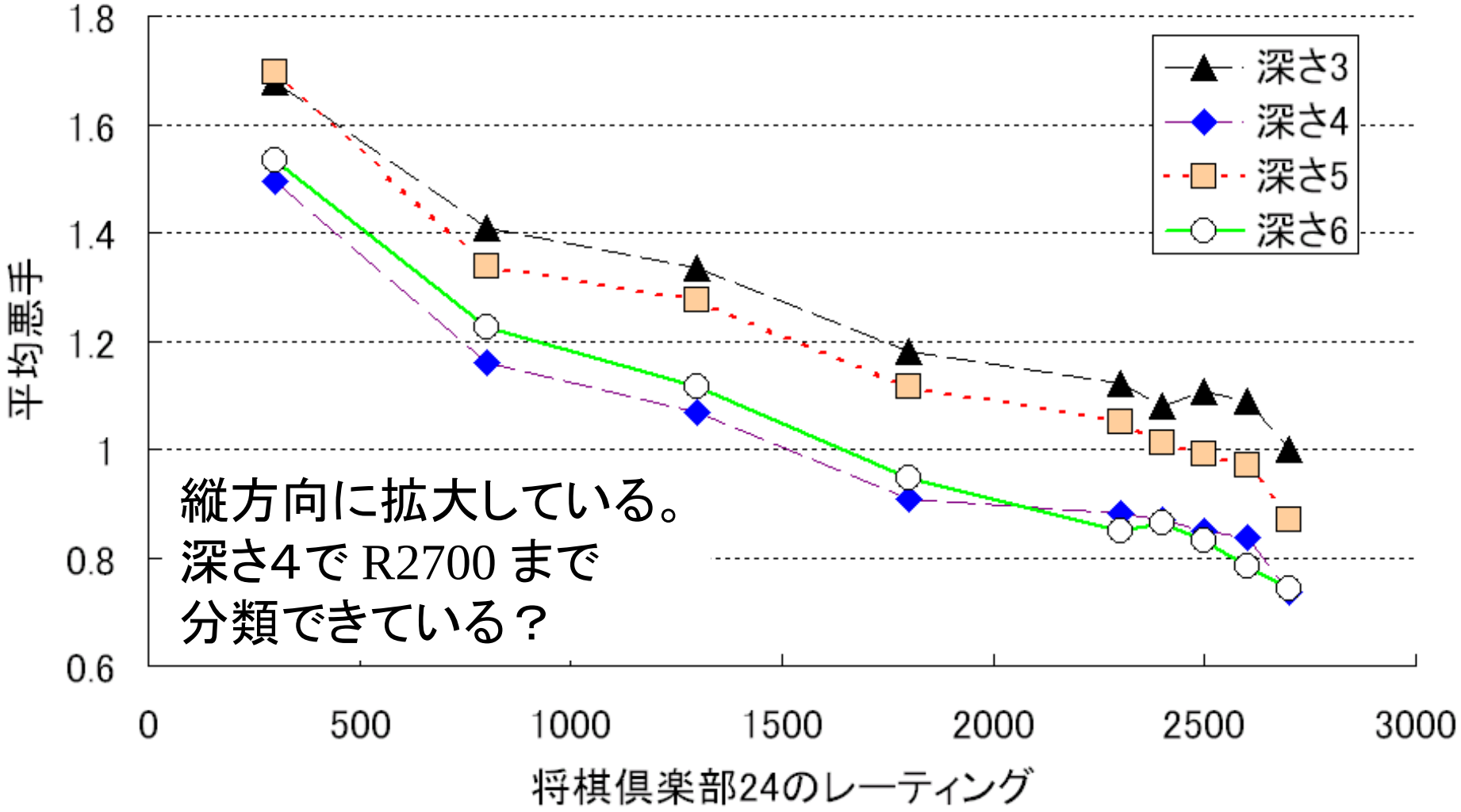
GPSFishでも同様の解析を行った

- GPSFish（2013年8月版）の深さ12固定
  - Bonanzaより1手深いが時間は1局8分と6分でGPSFishの方が短い
- Bonanzaの深さ11に319勝181敗
  - レーティングだと+99強い
  - より正確な解析ができると期待

深さ2、7、12の平均悪手（GPSFish）



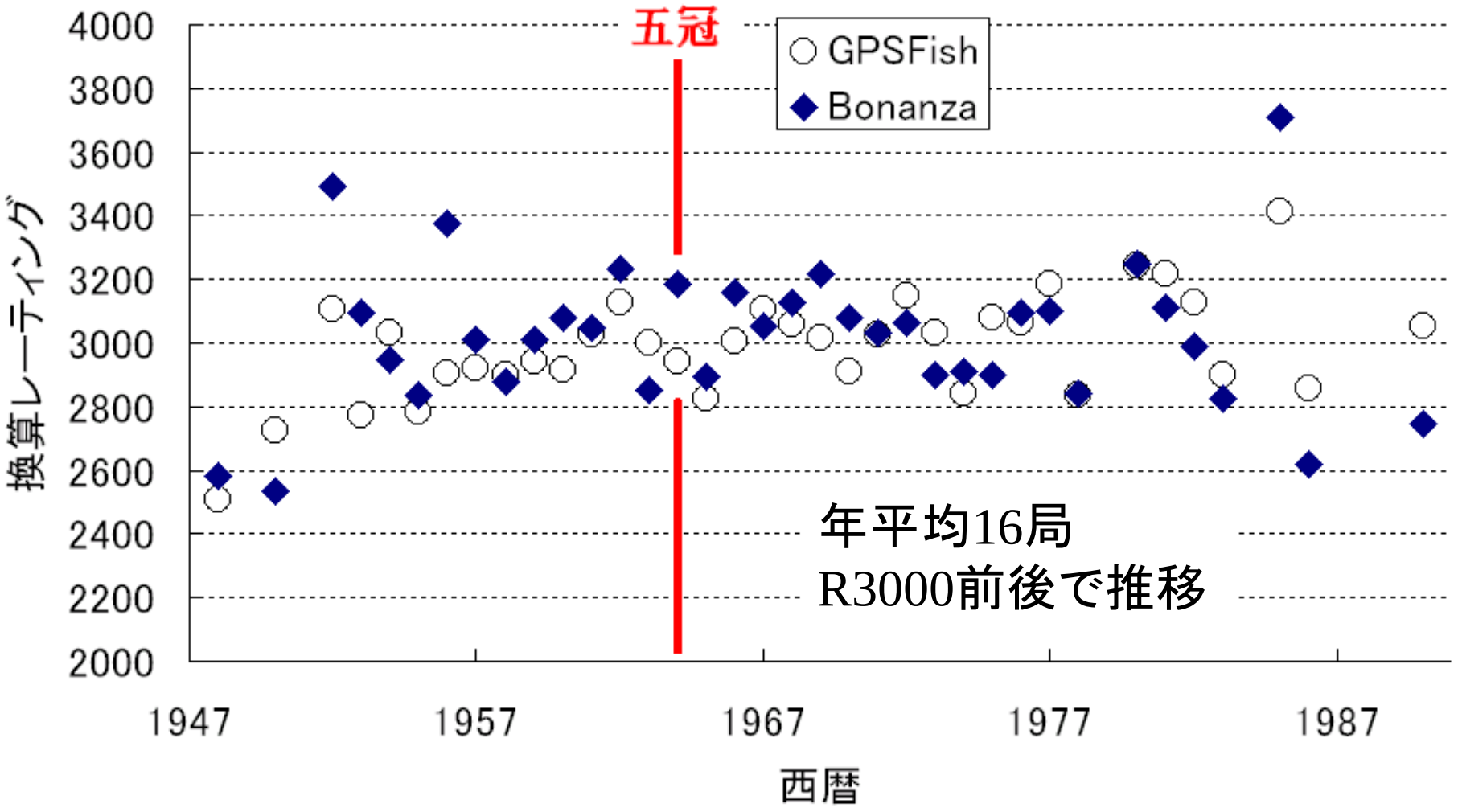
深さ3、4、5、6の平均悪手（GPSFish）



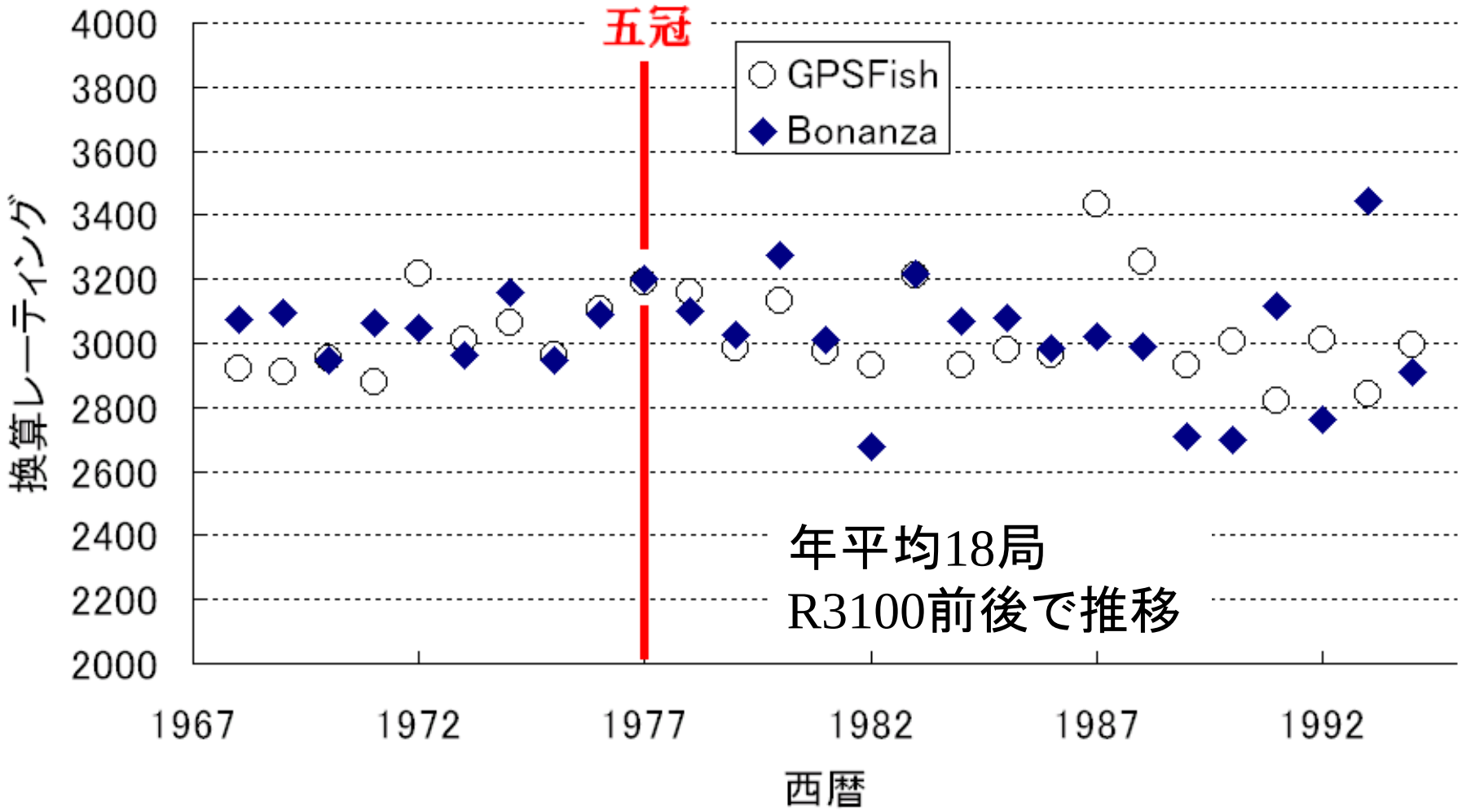
４人の名人のタイトル戦でのレーティング推移

- タイトル戦の平均悪手をレーティングに換算
- タイトル戦に限定したのはトップレベルの強さを発揮していた時期を比較したかったため
- 以下の4人
  - 大山康晴 15世名人
  - 中原誠　16世名人
  - 谷川浩司 17世名人
  - 羽生善治 19世名人

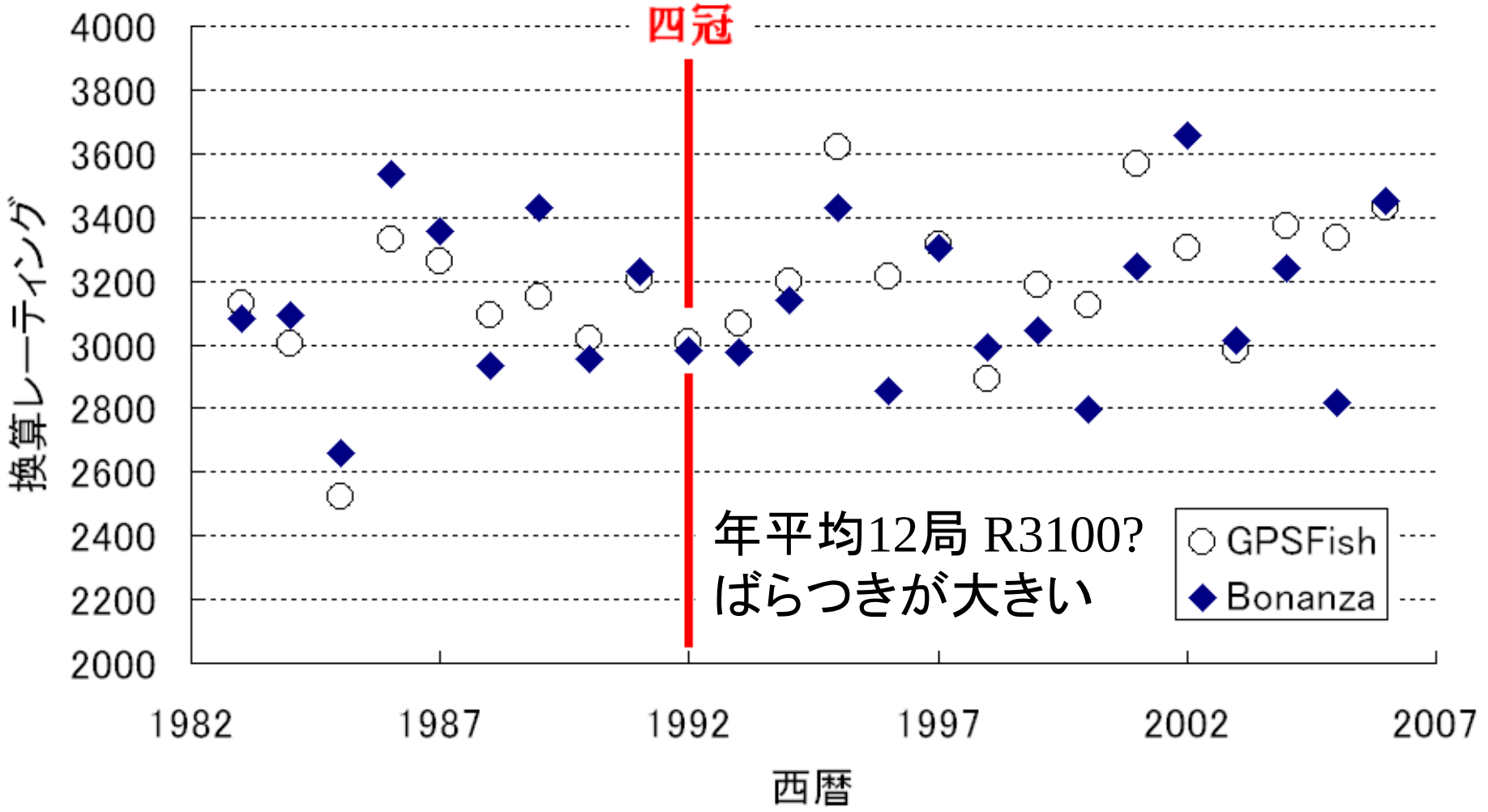
大山の換算レーティングの推移



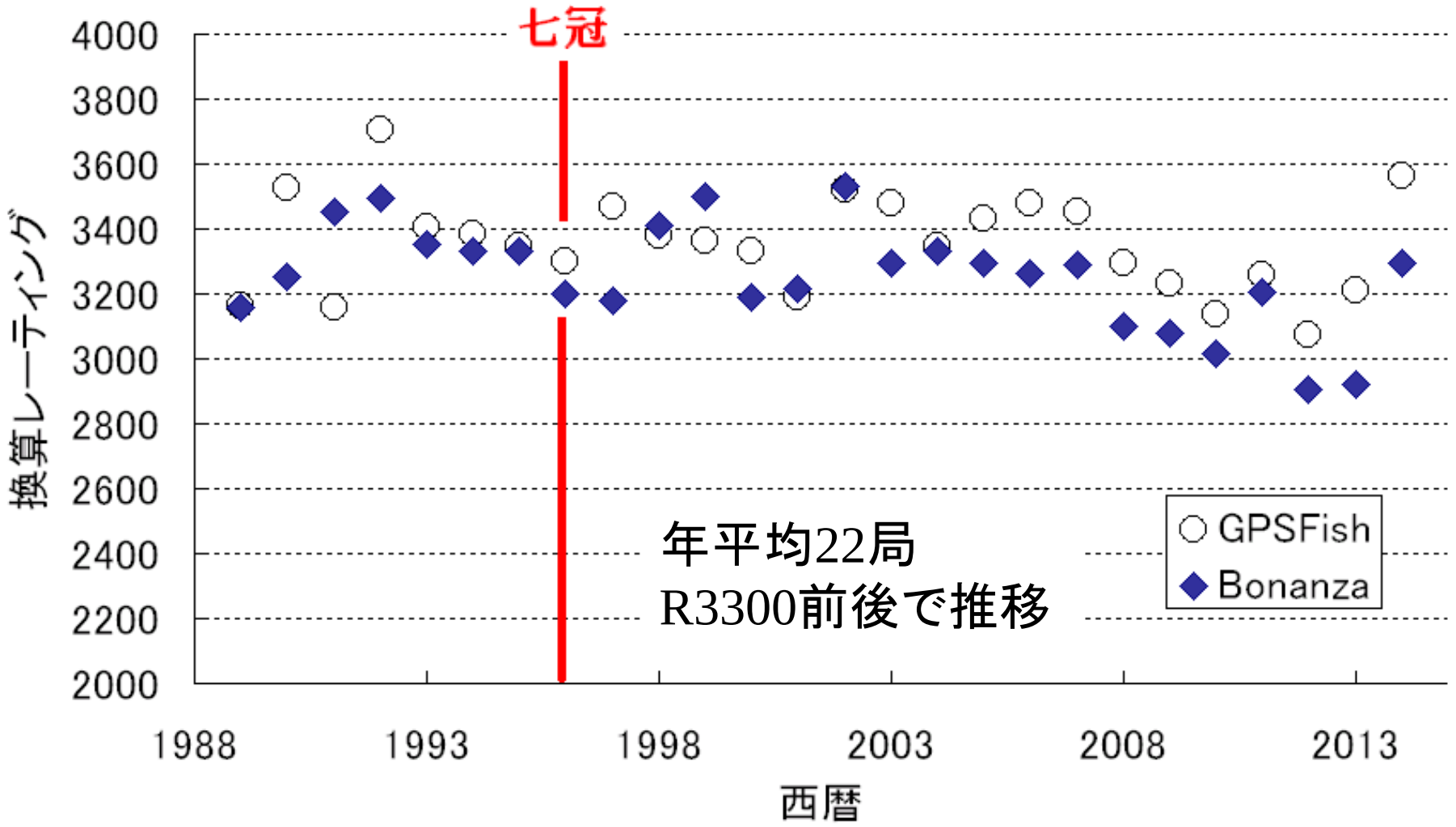
中原の換算レーティングの推移



谷川の換算レーティングの推移



羽生の換算レーティングの推移



羽生の点数が大山を上回る

- 羽生の1996年、大山の1964年、その前後2年の平均では羽生が227点上

GPSFishの方が変動が小さい

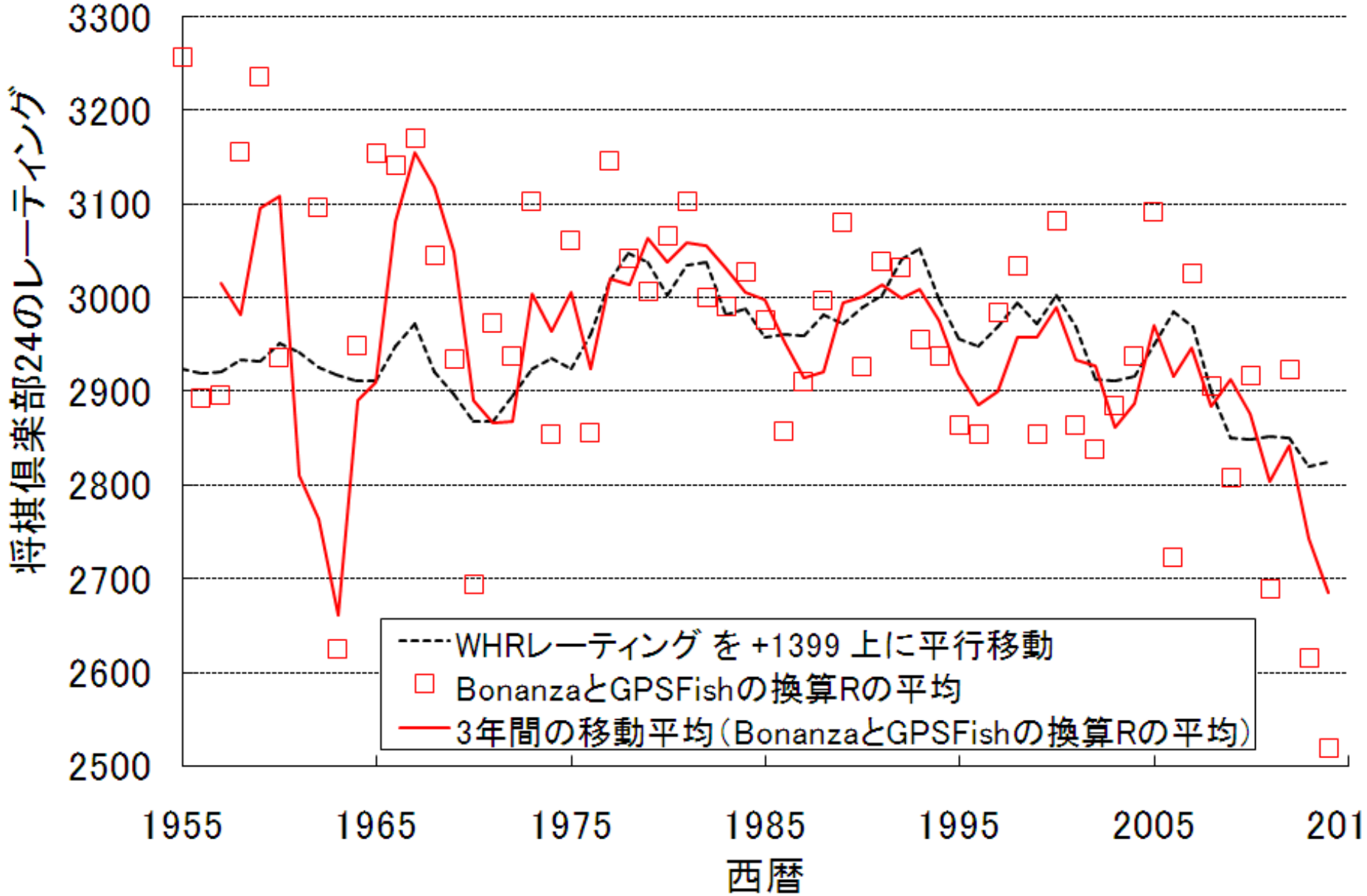
1年ごとの換算Rの変動の大きさ

	年間対局数	Bonanza	GPSFish
大山	16	242	170
中原	18	201	152
谷川	12	359	246
羽生	22	125	155
平均		**232**	**181**

加藤一二三のWHRレーティングと換算レーティングを比較する

- 棋譜は早指し戦を除く１４５５棋譜
  - タイトル戦以外も含む
- BonanzaとGPSFishの平均を取る
  - さらにその3年間の移動平均とWHRを比較

加藤のWHRと換算レーティング



- 勝敗の結果のみから計算したレーティングと棋譜の内容のみからら計算したレーティングがほぼ一致した

江戸時代の棋士の結果

	対局数	勝率	換算 R	(GPSFish)
初代大橋宗桂	8	0.875	**2555**	**2687**
本因坊算砂	16	0.250	**2611**	**2576**
初代伊藤宗看	30	0.667	**2510**	**2590**
六代大橋宗英	26	0.808	**2987**	**2731**
大橋柳雪	34	0.676	**2556**	**2399**
天野宗歩	45	0.756	**2758**	**2839**
八代伊藤宗印	48	0.792	**2776**	**2488**

家元制最強は宗英、を確認？

- 六代大橋宗英の換算Rが高い
- 『宗英は家元制で最強の名人とされる』
  - 関西将棋連盟のホームページより

タイトル戦（２日制）を基準とした１日制、NHK杯の差

- NHK杯の棋譜は両対局者のEloレーティングの合計が高いものから年間上位10局
- タイトル戦と同レベルの棋士が指した棋譜を集めた

	2 日制	1 日制の差	NHK 杯の差
渡辺明	3008	−246	−63
森内俊之	3027	−126	+6
羽生善治	3275	−79	−63
谷川浩司	3115	−72	−205
中原誠	3039	−41	−286
加藤一二三	3011	−110	−91
大山康晴	3036	−47	−296
升田幸三	3027	−356	−230
平均		**-103**	**-153**
(GPSFish)		**-108**	**-255**

２日制と１日制には明確に差がある

- ２日制に比べて
  - １日制は　100点低い
  - NHK杯は　200点低い
    - BonanzaとGPSFishの平均で
- 加藤一二三はNHK杯で91点低い、と小さい
  - 「秒読みの神様」を確認？

ソフトの換算レーティング

	対局数	持時間	換算 R	GPSFish	詰見逃し	率	敗北
NineDayFever	200	15 分	**3793**	**3813**	35	17.5	0
gpsfish_Xeon	11	6 時間	**3320**	**3820**	0	0.0	0
YssF_6t_x1	10	6 時間	**3633**	**3672**	0	0.0	0

- Floodgateの棋譜から
- 全般に換算Rが高すぎる
  - 15分のNDFより6時間のgpsfish_Xeonが強いはずだが分かってない
  - 詰見逃しはソフトには無関係らしい
  - ソフトの換算Rは信用できない？

NHK杯の羽生とdcsyhi(デクシ)の比較

	対局数	勝率	換算 R	GPSFish
羽生善治	86	0.826	**3213**	**3083**
dcsyhi	19	0.737	**2823**	**2698**

- dcsyhiは羽生ではなかった？
  - 400点差もある
  - dcsyhiの棋譜は羽生にしては弱すぎる
  - もしくはNHK杯の10分+30秒+考慮時間1分×10回と将棋倶楽部24の1分+30秒の差か？

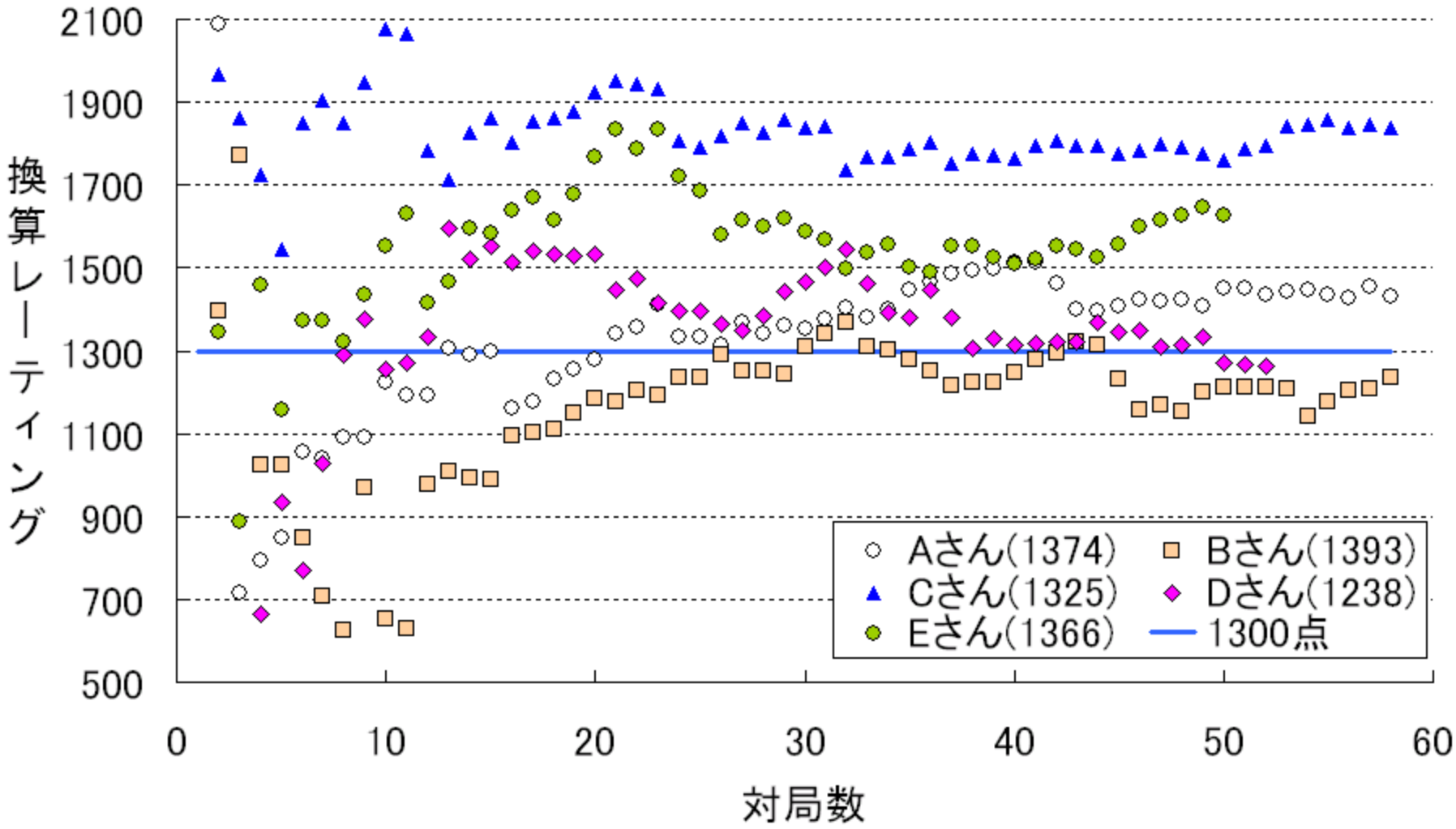
以降はやや否定的なデータです

アマチュアの人の換算Rを計算

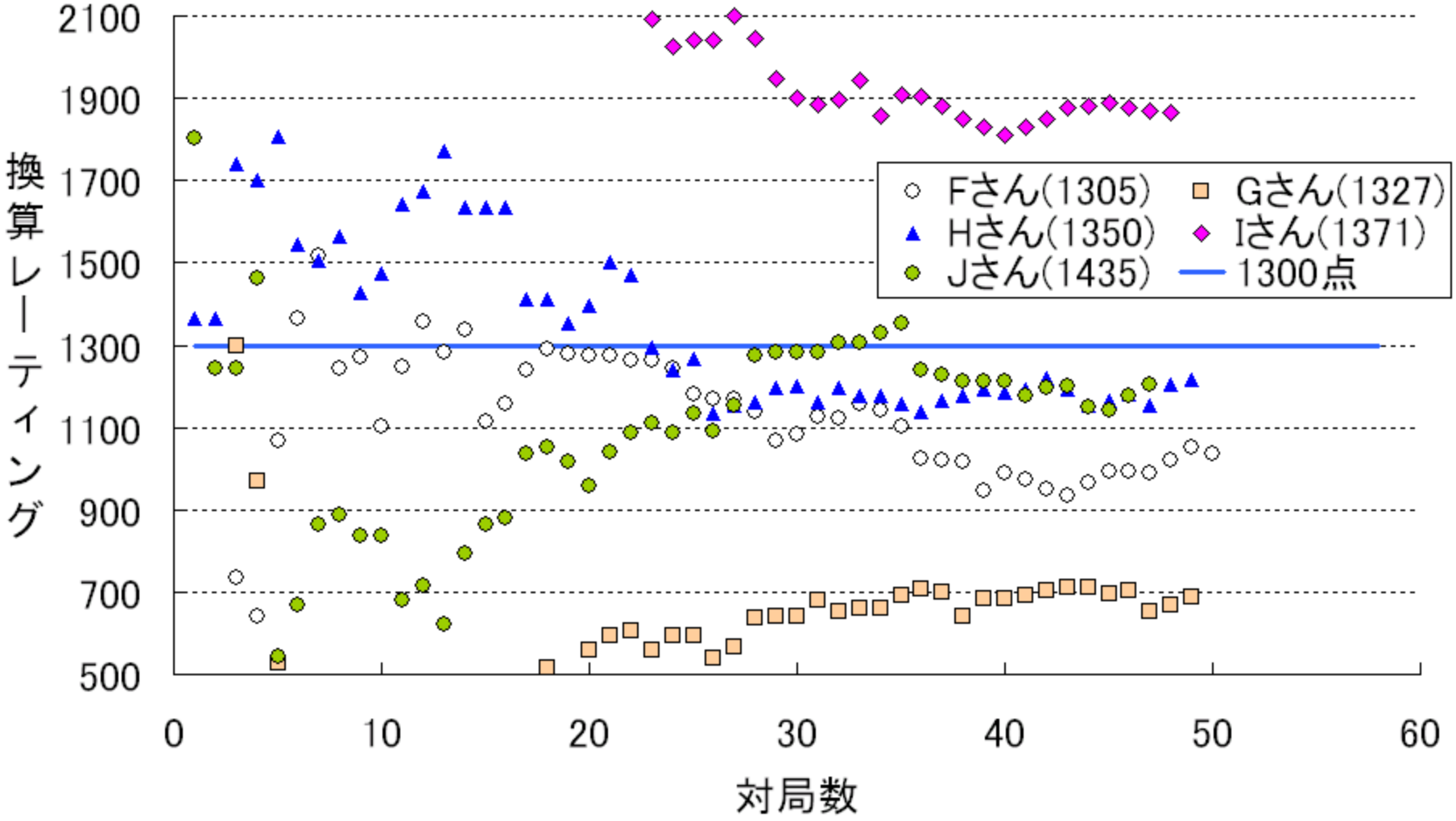
- 将棋倶楽部24の人達のレーティングを検証
  - 名前の後ろは平均レーティング
- 1300点(1350±30)で対局した棋譜を解析
- 1800点(1850±30)
- 2300点(2350±30)

- 対局日は関係なくランダムに並び替え
- 論文含め、換算Rは+50を足すのが正しい

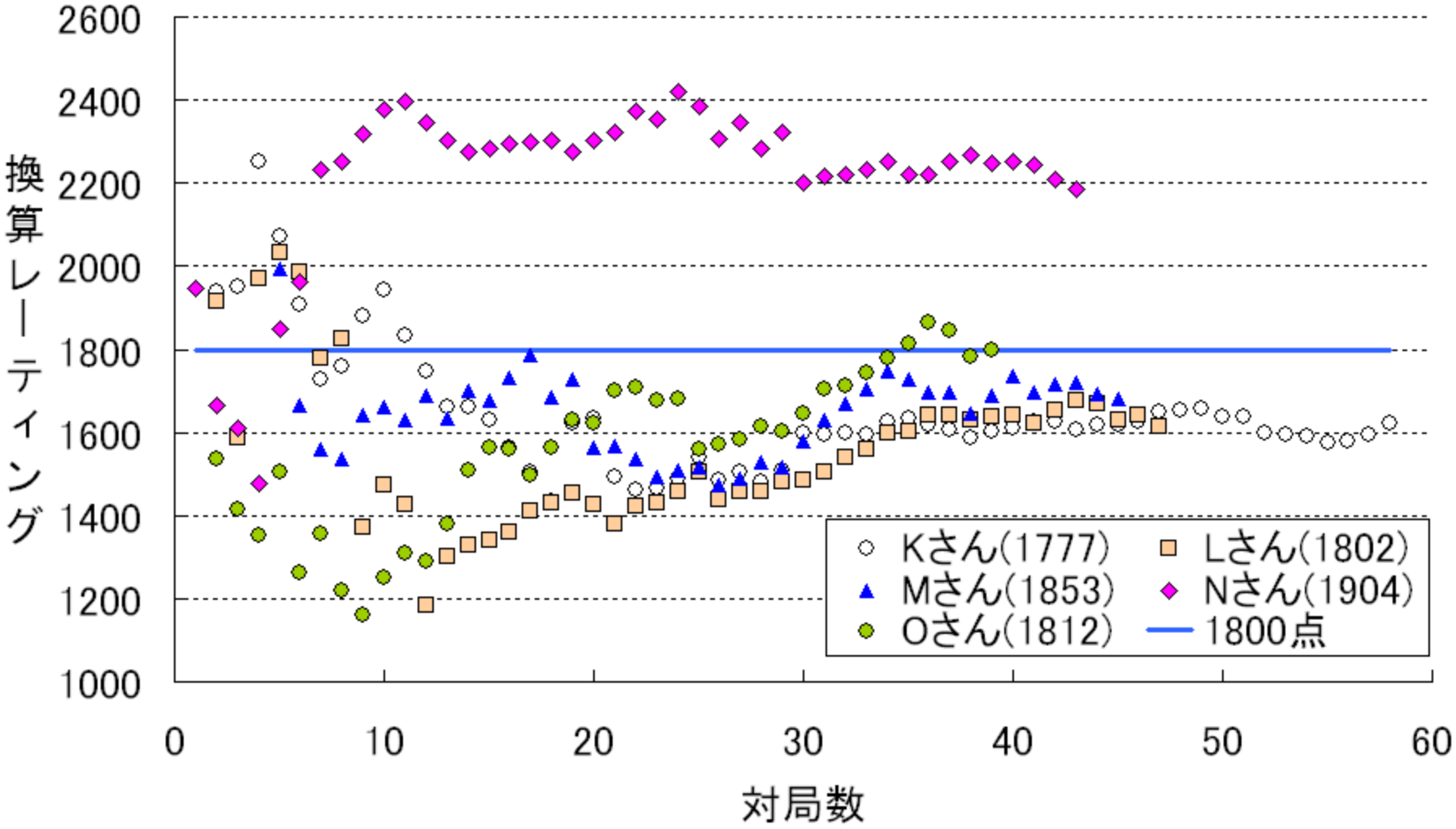
1300点の人達の換算R



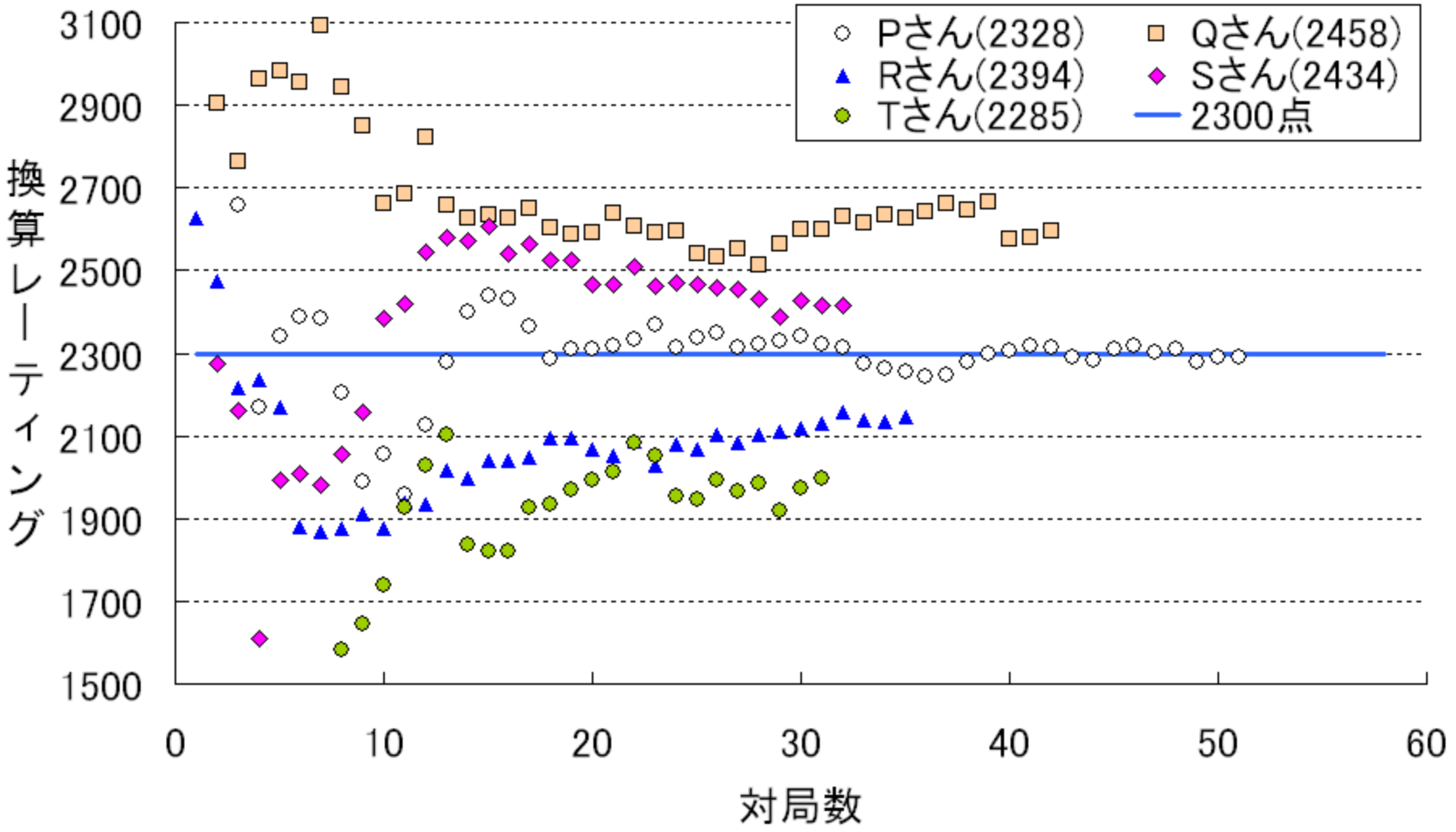
1300点の人達の換算R その2



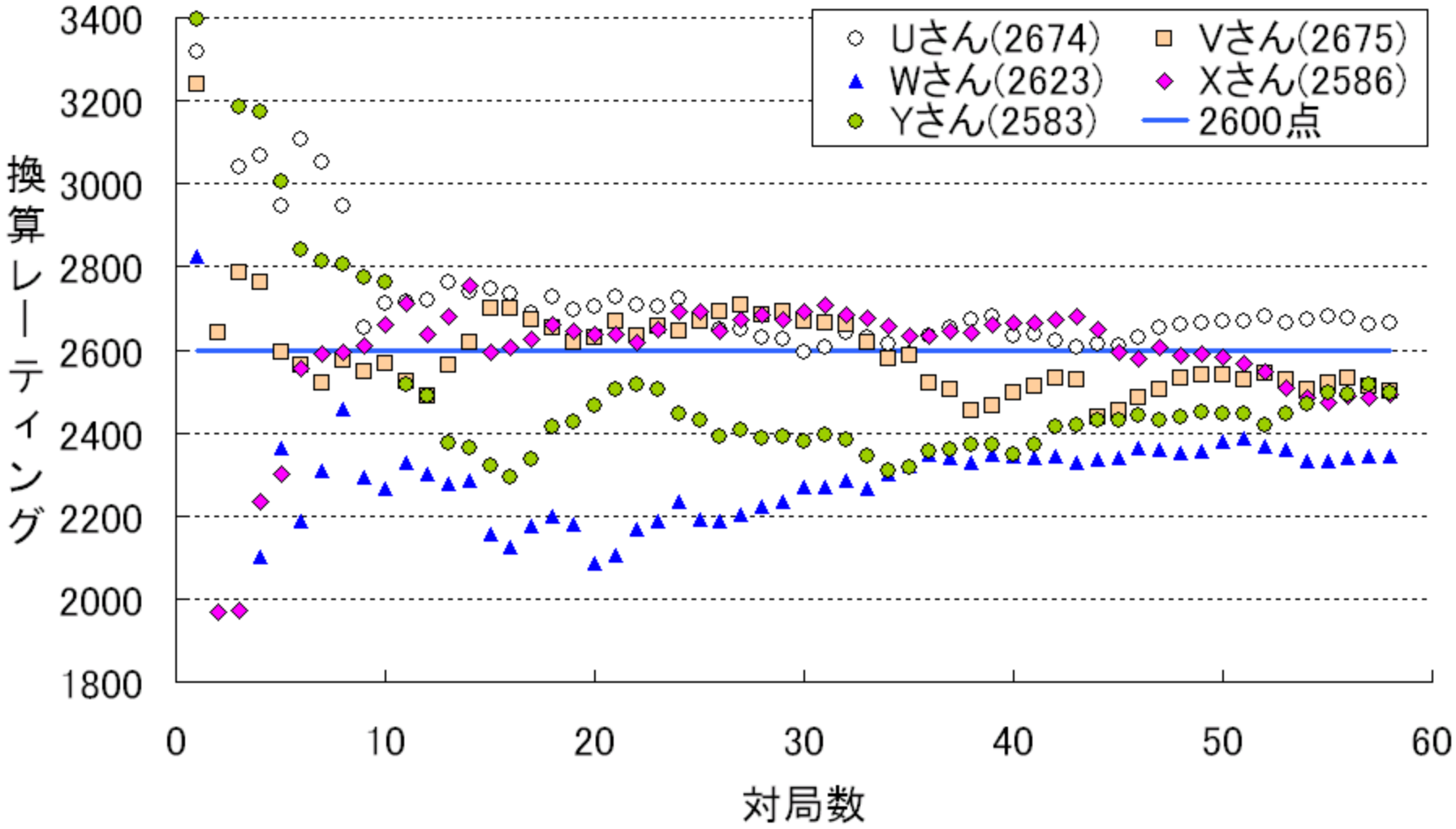
1800点の人達の換算R



2300点の人達の換算R



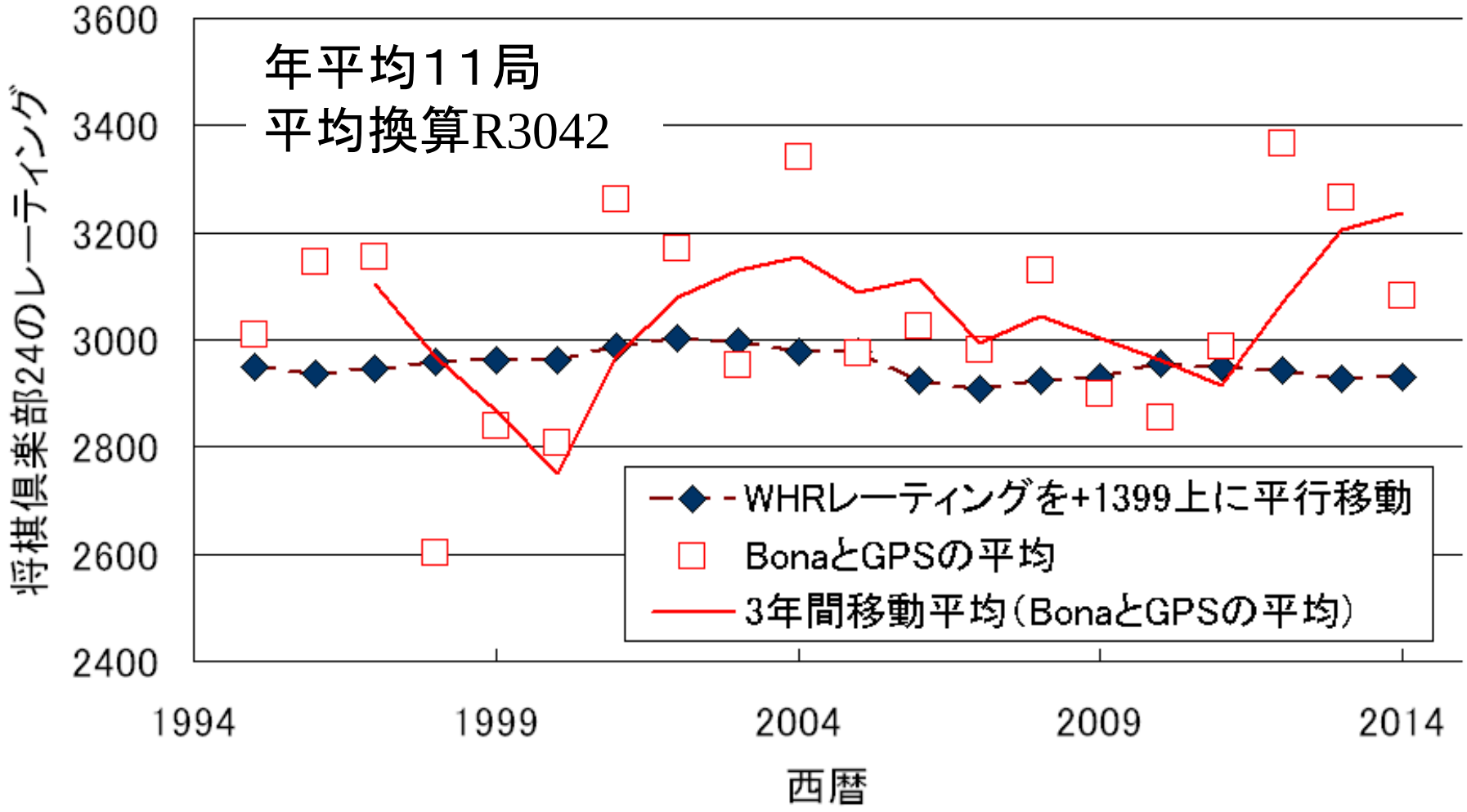
2600点の人達の換算R



かなりブレが大きい

- 棋風によって500点程度はずれる？
  - Rが高いほどぶれは減っている
- 持ち時間の差が大きい？（棋譜に情報なし）
  - 早指し ・・・ 1分＋1手30秒
  - 15分 ・・・ 15分＋1手1分
  - R1300は7割が早指し
  - R1800は9割が早指し
- 対局数は24局程度は必要か？

勝又さんの換算レーティング推移



勝又さんの換算レーティング

- WHRはほとんど動きがない
- 換算Rの平均は3042
- WHRの平均はR2952で90点差
  - 加藤一二三さんと同じだけ並行移動した場合
  - 女流を含めた平均はR2837(1438+1399)
- 換算Rはタイトル戦の中原16世名人に近い

神様のレーティングはR4600？

- Bonanza
  - rating = -3148y + 4620
- GPSFish
  - rating = -2560y + 4743
- 平均悪手 y=0 で 4620 と 4743
- このあたりが神様のレーティングか？

仮定に仮定を重ねた結果ですが

- 羽生の7冠時代は大山の全盛期より227点強い
- 20棋譜程度ですべてのプレイヤの棋力を推定できる
- 大橋宗英が家元制で最強なのを確認した
- 2日制の棋譜は1日制、NHK杯よりはっきりレベルが高い

今後の課題

- より予測が正確な指標の発見
  - 平均悪手と好手率、複雑さの組み合わせなど
- より正確で、多くの棋譜の入手
- 将棋倶楽部24のレーティングが正しいか検証
  - R2700以上が増えるなどインフレしている？
  - 持ち時間を統一して調べる
- より深い探索での解析
- より多くのプログラムでの解析